



ADVANCED ANALYTICAL TOOLS FOR THE INVESTIGATION OF LICIT AND ILLICIT DRUG RESIDUES IN WATER

Richard Thomas Dacre Northcote Bade

July 2016



Doctoral Supervisors

Prof. Dr. Félix Hernández Hernández

Prof. Dr. Juan Vicente Sancho Llopis

Prof. Dr. Pim de Voogt



Universitat Jaume I
Departament de Química Física i Analítica
Institut Universitari de Plaguicides i Aigües

**ADVANCED ANALYTICAL TOOLS FOR THE
INVESTIGATION OF LICIT AND ILLICIT DRUG
RESIDUES IN WATER**

Doctoral Thesis
Richard Thomas Dacre Northcote Bade
2016

El Dr. **Félix Hernández Hernández**, **Dr. Juan Vicente Sancho Llopis**, Catedráticos de Química Analítica, de la Universitat Jaume I de Castellón y **Prof. Dr. Pim de Voogt**, Catedrático de Química Ambiental, de la University of Amsterdam,

Certifican: que la Tesis Doctoral “Advanced Analytical Tools for the Investigation of Licit and Illicit Drug Residues in Water” ha sido desarrollada bajo su dirección, en el Instituto Universitario de Plaguicidas y Aguas, Departamento de Química Física y Analítica de la Universitat Jaume I de Castellón, por **Richard Thomas Dacre Northcote Bade**.

Lo que certificamos para los efectos oportunos en Castellón de la Plana, a de 14 Abril de 2016.

Fdo. Dr. Félix Hernández Hernández

Fdo. Dr. Juan Vicente Sancho Llopis

Fdo. Dr. Pim de Voogt

This thesis has been carried out, and will consequently be defended, in order to obtain the title “Doctor internacional”. The PhD candidate fulfilled the requirements for an international thesis as regulated within the framework of the Real Decreto 99/2011, “Normativa de los estudios de doctorado”, at the University Jaume I.

Richard Bade performed five secondments for a total of five months. The first secondment was carried out at Mario Negri Institute of Pharmacological Research in Milan, Italy from April 28-May 26 2014. The work entitled “Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using LC-LTQ-Orbitrap MS” has been performed under the supervision of Dr. Sara Castiglioni. The stay at this research centre enabled the candidate to gain knowledge in the use of an LTQ-Orbitrap for the analysis of environmental water samples. The second (and third) secondment was carried out at KWR Watercycle Research Centre, Nieuwegein, The Netherlands from August 18-September 12 2014 and February 2-21 2015. The work entitled “Evaluation of deconvolution tools for processing of high resolution mass spectrometry data in the screening analysis of wastewater” was carried out under the supervision of Erik Emke and Prof. Dr. Pim de Voogt. The stay at this research centre allowed the candidate to learn about new, advanced deconvolution tools to aid in suspect and non-target screening strategies of wastewater. The fourth secondment was carried out at the University of Saarland, Germany from 1-18 June 2015. The work entitled “*In vitro* study of 3-fluorophenmetrazine” was carried out under the supervision of Prof. Dr. Hans Maurer. The stay at this research centre allowed the candidate to learn how to perform *in vitro* studies, of particular importance for new psychoactive substances. The final secondment was made at the University of Antwerp, Belgium from 20 July-14 August 2015. The work “Evaluation of retention time prediction for different LC columns (C8, Biphenyl and HILIC) using LC-QTOF-MS” was carried out under the supervision of Dr. Alexander van Nuijs and Prof. Dr. Adrian Covaci. This stay allowed the candidate to test his (previously developed) retention time prediction method on different columns and using different compounds for optimisation.

The work performed in this thesis was part of the activities described in the projects (Research Group of Excellence PROMETEO II/2014/023; and ISIC 2012/ 016) of the Generalitat Valenciana, the Spanish Ministry of Economy and Competitiveness (Project ref CTQ2015-65603) and the European Union-International Training Network SEWPROF (Marie Curie — PEOPLE Grant #317205).

Prior to the defence of the Doctoral thesis, this work was independently evaluated by two external reviewers, Dr. B. Kasprzyk-Hordern (Department of Chemistry, University of Bath, United Kingdom) and Dr. J. Mueller (National Research Centre for Environmental Toxicology (Entox), The University of Queensland, Australia).

For my family

Summary

This doctoral thesis covers the use of advanced analytical techniques (UHPLC-QTOF, UHPLC-QqQ-MS/MS and LC-LTQ-Orbitrap) and computational software (Artificial Neural Networks, Sieve, MsXelerator) from the prediction of chromatographic retention time to the qualitative screening and quantification of pharmaceuticals and illicit drugs (PIDs) in wastewater (WW) and surface water (SW). It is divided into three sections and consists of six scientific articles and one review article (**Scientific Article 1**).

The first section consists of three research articles and deals with the use of high resolution mass spectrometry (HRMS) in the screening and occurrence of PIDs in environmental waters of three countries: Colombia, Italy and Spain. The use of different HRMS instruments (QTOF and LTQ-Orbitrap) combined with in-house and online databases of up to thousands of compounds enabled numerous compounds to be detected and identified. A continuously-updated in-house database of more than 1000 PIDs and metabolites was made incorporating their exact mass, and also adduct formation, fragment ions and retention time when a standard was available. This entire database was used for the first work, investigating influent and effluent WW as well as SW samples from Bogotá, Colombia, leading to the identification of 40 compounds across all samples (**Scientific Article 2**). A subset of approximately 100 of the most commonly identified compounds was used in the next work, looking at the use of two HRMS instruments (QTOF and Orbitrap) in the suspect screening of WW and SW samples from Italy and Spain, with up to 28 compounds at least detected (**Scientific Article 3**). Finally, a non-target screening approach was made using two advanced deconvolution software and incorporating two databases: one of 250 PIDs and the ChemSpider database, with more than 20 compounds being tentatively identified (**Scientific Article 4**).

The second section consists of two research articles and covers the development and utilisation of two retention time prediction approaches in the identification of PIDs in waste and environmental water samples. Initially, a very simple approach was made, using only $\log K_{ow}$ and a combined in-house and Waters database of 595 reference standards. This led to a prediction accuracy of approximately 70-80% within a retention time window of ± 2 minutes. A comparison was made to see whether reducing the mass extraction window incrementally from 50mDa to 5mDa would result in more peaks being removed than the prediction method.

It was found that even at the narrowest window investigated, this prediction method resulted in almost one quarter of all chromatographic peaks able to be removed as false positives (**Scientific Article 5**). While this simple approach showed itself quite useful, a more precise method was hoped to aid in the identification of transformation products for which no reference standards are available. Artificial neural networks were thus investigated and found to be much more precise, with up to 95% of all compounds being within a ± 2 minute window, leading to the tentative identification of ten metabolites and transformation products and several other compounds for which standards were unavailable (**Scientific Article 6**).

Finally, the third section consists of one research article on the quantitative analysis of new psychoactive substances (NPS) using UHPLC-QqQ-MS/MS. Ten NPS were chosen, based on their current (ab)use and prevalence in toxicological and wastewater-based epidemiological approaches. This method was based on offline solid phase extraction (SPE) using mixed cation exchange (MCX) cartridges due to the basic nature of the NPS. A study was also made on the stability of the compounds in influent wastewater (both at natural pH and pH=2), with the compounds maintaining their stability for up to 7 days in acidified influent wastewater. This method was fully validated and tested on samples from throughout Europe – Utrecht, The Netherlands; Oslo, Norway; Bristol, United Kingdom; Milan, Italy; Copenhagen, Denmark; Brussels, Belgium, Zurich, Switzerland and Castellon, Spain – with mephedrone, methylone and methylenedioxypropylone (MDPV) able to be identified (**Scientific Article 7**).

Resumen

Esta tesis doctoral abarca el uso de técnicas avanzadas de análisis (HPLC-QTOF, HPLC-QqQ-MS/MS y LC-LTQ-Orbitrap) y software computacional (Artificial Neural Networks, Sieve, MsXelerator) desde la predicción de tiempo de retención cromatográfico a la detección cualitativa y cuantitativa de productos farmacéuticos y drogas ilícitas (PIDs) en aguas residuales (WW) y superficiales (SW). Se divide en tres secciones y consta de seis artículos científicos y un artículo de revisión.

La primera sección consta de tres artículos de investigación y se ocupa de la utilización de la espectrometría de masas de alta resolución (HRMS) en el “screening” de PIDs en aguas ambientales de tres países: Colombia, Italia y España. El uso de diferentes instrumentos de HRMS (QTOF y LTQ-Orbitrap) combinado tanto con bases de datos internas como en línea de hasta miles de compuestos ha permitido la detección e identificación de numerosos compuestos. Se ha desarrollado una base de datos que se actualiza de manera continua, con más de 1000 PIDs y metabolitos que incluye su masa exacta, así como la formación de aductos, los iones fragmento así como el tiempo de retención cuando se contaba con un patrón. Esta base de datos se utilizó para el primer artículo, la investigación de influentes y efluentes de WW, así como muestras de SW de Bogotá, Colombia, que permitió la identificación de hasta 40 compuestos en todas las muestras (**Scientific Article 2**). Una base de datos reducida, conteniendo un subconjunto de aproximadamente 100 de los compuestos más comúnmente identificados, se utilizó en el siguiente artículo, evaluando la utilización de dos instrumentos HRMS (QTOF y Orbitrap) para el “suspect screening” de muestras WW y SW de Italia y España, con un máximo de 28 compuestos detectados (**Scientific Article 3**). Por último, un método combinado de “suspect” y “non-target” se ha llevado a cabo mediante dos programas de deconvolución avanzada y la incorporación de una pequeña base de datos de 250 PIDs para “suspect screening” y la base de datos generalista ChemSpider para el “non-target screening”, con casi 20 compuestos identificados tentativamente (**Scientific Article 4**).

La segunda sección consta de dos artículos de investigación que incluyen el desarrollo y la utilización de dos enfoques distintos para la predicción del tiempo de retención cromatográfico en la identificación de PIDs en muestras de aguas residuales y ambientales. Inicialmente, se realizó un enfoque muy simple, utilizando sólo $\log K_{ow}$ pero combinando una base de datos interna y otra de Waters, lo que permitió emplear los datos de 595 patrones de referencia. Esto condujo a una exactitud

en la predicción de aproximadamente el 70-80% de todos los compuestos dentro de una ventana de tiempo de retención de ± 2 minutos. Se llevó a cabo una comparación reduciendo la ventana de extracción de masa desde 50mDa a 5mDa para evaluar si conllevaría la eliminación de más picos cromatográficos que el método de predicción de tiempo de retención. Se observó que incluso con la ventana más estrecha investigada, este método de predicción de tiempo de retención permitió eliminar casi un cuarto de todos los picos cromatográficos como falsos positivos (**Scientific Article 5**). Si bien este enfoque simple se mostró bastante útil, se requiere un método más exacto para ayudar en la identificación de los productos de transformación para los que no hay patrones de referencia disponibles. Se investigó el uso de las redes neuronales artificiales (ANNs), resultando ser mucho más precisas, con hasta un 95% de todos los compuestos dentro de una ventana ± 2 minutos, dando lugar a la identificación tentativa de diez metabolitos y productos de transformación así como otros compuestos para los que los patrones de referencia no estaban disponibles (**Scientific Article 6**).

Finalmente, la tercera sección consiste en un artículo de investigación para el análisis cuantitativo de nuevas sustancias psicoactivas (NPS) utilizando HPLC-QqQ-MS/MS. Diez NPS fueron elegidos, en función de su (ab)uso actual y prevalencia tanto toxicológica como en epidemiología de aguas residuales. Este método se basa en un tratamiento previo de extracción en fase sólida (SPE), usando cartuchos de modo mixto de intercambio catiónico (MCX), debido a la basicidad de los NPS investigados. También se realizó un estudio sobre la estabilidad de los compuestos en agua residual influente a pH natural y a pH 2, resultando estables durante un máximo de 7 días en el agua residual acidificada. Este método fue totalmente validado y evaluado en muestras procedentes de toda Europa - Utrecht, Países Bajos; Oslo, Noruega; Bristol, Reino Unido; Milan, Italia; Copenhague, Dinamarca; Bruselas, Bélgica, Zurich, Suiza y Castellón, España – identificando en las muestras mefedrona, metilona y methylenedioxyprovalerone (MDPV) (**Scientific Article 7**)

Table of Contents

List of Abbreviations	ix
Objectives and Working Plan	3
Objetivos y Plan de Trabajo	9
1. General Introduction	
1.1 Emerging Contaminants in Water	17
1.1.1 Wastewater-based Epidemiology	18
1.2 Analytical Strategies	20
1.2.1 Liquid Chromatography coupled to Mass Spectrometry	20
1.2.2 Matrix Effects	26
1.2.3 Target, Suspect and Non-target Screening	28
1.3 Scientific Article 1	33
Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry	
<i>Trends in Analytical Chemistry</i> (2014) 63: 140-157	
1.4 Marie Curie SEWPROF ITN	85
1.5 Scientific Articles Included in this Thesis	87
2. Qualitative Screening with Liquid Chromatography coupled with High Resolution Mass Spectrometry	
2.1 Introduction	91
2.2 Scientific articles	95
2.2.1 Scientific Article 2	97
LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia	
<i>Analytical and Bioanalytical Chemistry</i> (2015) 407: 6405-6416	

2.2.2	Scientific Article 3	121
	Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS	
	<i>Analytical and Bioanalytical Chemistry</i> (2015) 407:8979-8988	
2.2.3	Scientific Article 4	143
	Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools	
	<i>Science of the Total Environment</i> (2016) UNDER REVISION	
2.3	Discussion of Results	167
3.	Retention time prediction in the Identification of Emerging Contaminants in the Aquatic Environment	
3.1	Introduction	175
3.2	Scientific Articles	181
	3.2.1 Scientific Article 5	183
	Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water	
	<i>Talanta</i> (2015) 139: 143-149	
	3.2.2 Scientific Article 6	205
	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	
	<i>Science of the Total Environment</i> (2015) 538: 934-941	
3.3	Discussion of Results	227

4	Quantitative determination of selected New Psychoactive Substances in environmental water samples	
4.1	Introduction	237
4.2	Scientific Articles	243
4.2.1	Scientific Article 7	245
	Ultra-high performance liquid chromatography-tandem mass spectrometry determination of selected new psychoactive substances including synthetic cathinones and phenethylamines in influent wastewater of eight European cities	
	SUBMITTED FOR PUBLICATION	
4.3	Discussion of Results	277
5	References	289
6	Conclusions/Conclusiones	
6.1	Conclusions	307
6.2	Conclusiones	309
6.3	Suggestions for Future Research	313
6.4	Sugerencias para Futuras Investigaciones	315
	Curriculum Vitae	317
	Acknowledgements	325

List of Abbreviations

25-B-NBOMe: 2-(4-bromo-2,5- dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine

25-C-NBOMe: 2-(4-chloro-2,5- dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine

25-I-NBOMe: 2-(4-iodo-2,5- dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine

ANN: Artificial neural networks

CID: Collision induced dissociation

DDA: Data dependent acquisition

DDT: dichlorodiphenyltrichloroethane

EC: Environmental contaminant

ESI: Electrospray ionisation

FA: Formic acid

FT-ICR: Fourier-transformed ion cyclotron resonance mass analyser

H₂O: Water

HILIC: Hydrophobic interaction liquid chromatography

HLB: Hydrophilic-lipophilic-balanced

HRMS: High resolution mass spectrometry

IT: Ion trap mass analyser

ITN: International training network

KWR: KWR Watercycle Research Institute

LC: Liquid chromatography

LC-MS: Liquid chromatography coupled to mass spectrometry

LC-MS/MS: Liquid chromatography coupled to tandem mass spectrometry

log_{K_{ow}}: Octanol-water partition coefficient

LSER: Linear solvent energy relationship

LTQ: Linear ion trap mass analyser

MCX: Mixed-mode cation exchange

MDMA: 3,4-methylenedioxyamphetamine, also known as ecstasy

MDPV: methylenedioxypropylone

MeOH: Methanol

MN: Mario Negri Institute of Pharmacological Research

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

m/z : mass to charge ratio

NH₄Ac: Ammonium acetate

NP: Normal phase

NPS: New psychoactive substances

nw-XIC: Narrow window extracted ion chromatogram

OECD: Organisation for Economic Co-operation and Development

PCB: Polychlorinated biphenyls

PID: Pharmaceuticals and illicit drugs

ppb: Parts per billion

ppt: Parts per trillion

q: non-mass resolving quadrupole, acting as a collision cell

Q: Quadrupole mass analyser

QqQ: Triple quadrupole mass analyser

QSRR: Quantitative structure-retention relationships

QTOF: Quadrupole time-of-flight mass analyser

RF: Random forest

RP: Reversed phase

SIM: Selected ion monitoring

SMILES: Simplified molecular-input line-entry system

SPE: Solid phase extraction

SRM: Selected reaction monitoring

SVM: Support vector machine

t_R : Retention time

TIC: Total ion chromatogram

TOF: Time-of flight mass analyser

TP: Transformation product

UHPLC: Ultra high performance liquid chromatography

WBE: Wastewater based epidemiology

WWTP: Wastewater treatment plant

XIC: Extracted ion chromatogram

OBJECTIVES AND WORKING PLAN

OBJETIVOS Y PLAN DE TRABAJO

Objectives

The main objective of this doctoral thesis was to investigate licit and illicit drugs and metabolites in waste and surface water using advanced analytical techniques. To reach this aim, retention time prediction, suspect, non-target and target screening strategies have been applied. Different instrumental configurations have been investigated, including low resolution tandem mass spectrometry (MS) using triple quadrupole (QqQ) and high resolution MS using quadrupole time-of-flight (QTOF) and linear ion trap Orbitrap (LTQ Orbitrap) mass analysers.

In order to reach this general objective, several specific objectives have been established:

- 1) Application of a qualitative analytical methodology based on ultra-high performance liquid chromatography (UHPLC) coupled to QTOF MS in MS^E mode, allowing rapid wide-scope screening for a large number of compounds based on a large in-house database. This methodology was to be applied to a number of water samples (influent and effluent wastewater and surface water) from different geographic origins and countries (Colombia, Italy and Spain).
- 2) Development of a simple retention time prediction approach that can easily be applied during the suspect screening of pharmaceuticals and illicit drugs to aid in their identification. A comparison with mass extraction window width was also to be tested.
- 3) Development of a second, more complex and precise retention time prediction approach that could be applied in the identification of metabolites and transformation products for which standards are unavailable.
- 4) Development and application of a combined suspect and non-target screening workflow using advanced deconvolution software following injection in LTQ-Orbitrap.
- 5) Development of an analytical strategy for quantification of new psychoactive substances (NPS) in influent wastewater based on offline solid phase extraction (SPE) followed by LC-MS/MS measurement with a triple quadrupole mass analyser. A comparative direct injection method was also performed. Target compounds were selected based on consumption and prevalence in previous toxicological and wastewater samples.

Working Plan

This entire doctorate was supported by the Marie Curie funded International Training Network, SEWPROF (more in **Chapter 1**), which made it possible to perform research stays at several international institutes.

The working plan followed in order to achieve the objectives outlined above was as follows:

Qualitative Screening with Liquid Chromatography coupled with High Resolution Mass Spectrometry

The work in this section was performed at three different institutes: University Jaume I, Mario Negri Institute for Pharmacological Research and KWR Watercycle Research Institute and the specific working plans within each institute is summarised below:

Working Plan for work performed at University Jaume I (Spain)

1. Continued upkeep of an in-house database developed by our group for the qualitative screening of pharmaceuticals and illicit drugs. This meant including (at least) the name and molecular mass, with the exact mass of fragment ions added when known.
2. Manual collection of surface water samples from different areas in the Valencian region. This entailed physically collecting samples in high density polyurethane bottles, transporting them to the laboratory and storing them at 4°C.
3. Sample treatment of wastewater and surface water samples following previously developed solid phase extraction techniques employed in our laboratory.
4. Injection of samples in UHPLC-QTOF in MS^E mode, thereby allowing simultaneous collection of low and high energy full-spectrum accurate mass data.
5. Data analysis on specialised software, Chromalynx (Within MassLynx, version 4.1, Waters Corporation) which facilitates wide-scope screening analyses by automatically extracting the exact mass of a compound at low energy (and high energy in the cases of fragment ions) in the form of an extracted ion chromatogram (XIC).
6. Further data analysis performed on the raw data for parent compounds and fragment ions to ensure no compounds are erroneously not detected or identified.

Working Plan for work performed at Mario Negri Institute for Pharmacological Research (Italy)

1. Initial training with LTQ-Orbitrap based on theory and practice.
2. Injection of samples in the LTQ-Orbitrap in data-dependent acquisition (DDA) mode, meaning that when a peak passed a certain pre-set threshold, MS/MS scans were made to acquire fragment ions.
3. Data analysis on specialised software (TraceFinder EFS, Thermo Scientific) based on the same database used at University Jaume I where “positive” findings are based on the exact mass of the protonated molecule together with isotopic pattern, fragment ions and a library search (the latter two only when available).
4. For closer examination of “positive” compounds, the samples were reinjected in the LTQ-Orbitrap in MS/MS mode, based on a target list of the “positive” compounds to give information on product ions.
5. Further data analysis performed on the raw data (in QualBrowser, Thermo Scientific) for parent compounds and fragment ions to ensure no compounds are erroneously not detected or identified.

Working Plan for work performed at KWR Watercycle Research Institute (The Netherlands)

1. Injection of samples in the LTQ-Orbitrap in data-dependent acquisition (DDA) mode, meaning that when a peak passed a certain pre-set threshold, MS/MS scans were made to acquire fragment ions.
2. Data analysis on specialised software for non-target (MsXelerator (MsMetrix) and Sieve 2.1 (Thermo Scientific)) analysis.
3. A hidden target screening was made after initial peak-picking parameters were set and the same database as in previous studies was used to detect compounds within the samples.
4. Additionally, following initial peak-picking, the ChemSpider database was used to detect compounds.

5. All samples were reinjected in MS/MS mode in the LTQ-Orbitrap based on a target list from the combined detected compounds found by both suspect and non-target analysis, to help initiate fragmentation.
6. Analysis of samples within QualBrowser (Thermo Scientific) was made to compare fragment ions observed experimentally with those found in literature or in online mass spectral databases for tentative identification.

Retention Time Prediction

Retention Time Prediction based on $\log K_{ow}$

1. Bibliographic revision of the current trends and means of retention time prediction.
2. Development of database based on reference standards in our laboratory and a Waters database.
3. SMILES codes found for all compounds in the database.
4. Predicted $\log K_{ow}$ values found using VCCLabs freeware, experimental values found in literature and online databases.
5. Linear regression made between $\log K_{ow}$ values (both experimental and predicted) and retention time; different subsets based on chemical family also made to improve correlation.
6. Investigation and comparison of the impact of a $\log K_{ow}$ predicted retention time and narrower extraction windows in the removal of false positive peaks for the detection of emerging contaminants (with reference standards available) in environmental waters.

Retention Time Prediction based on Artificial Neural Networks (in collaboration with Leon Barron, King's College London)

1. Bibliographic revision of current means of retention time prediction based on artificial neural networks (ANN).
2. Additional descriptors required for ANN approach found on Parameter Client (VCC Labs)
3. Optimisation of ANN method based on implementing different training:verification:blind set ratios, as well as different architectures and model types.
4. Investigation of the impact of ANN-based predicted retention times on the exclusion of false positive peaks for the detection of emerging contaminants (with reference standards available) in environmental waters.

5. Investigation of ANN-based predicted retention times in the tentative identification of metabolites and transformation products for which reference standards are unavailable

Quantitative Determination of Selected New Psychoactive Substances in Influent

Wastewater Samples

1. Selection of target new psychoactive substances (NPS) based on literature revision
2. Optimisation of MS and MS/MS parameters by injection of a mixture of all compounds in solvent.
 - i. Acquisition of full-scan MS spectra to establish the most appropriate ionisation mode and cone voltage of the precursor ion.
 - ii. Improve formation of the protonated molecule by using additives in the mobile phase (i.e. formic acid and ammonium acetate).
 - iii. Following optimisation of cone voltage of the precursor ion, collision energy is then optimised in order to select the most appropriate product ions, taking into account sensitivity and specificity (i.e. avoid non-specific transitions such as loss of water).
 - iv. Selection of at least two MS/MS transitions per compound: one for quantification and at least one other for confirmation.
3. Selection of chromatographic conditions by injecting standard solutions, with mobile phase altered using additives to improve the chromatographic peak shape and separation.
4. Study of the SPE process (comparing different SPE cartridges to gain an idea of the best for the study) using blank samples spiked before and after the SPE step.
5. Study the possibility of direct injection for this method.
6. Evaluate filtration recoveries for both vacuum and syringe filters.
7. Evaluate stability of all compounds at different temperatures (-20°C, 4°C and 20°C) for up to 7 days (30 days at -20°C).
8. Validation of the optimised methods by evaluating linearity, accuracy and precision from recovery experiments at different concentration levels. Estimation of the limits of quantification and detection of all studied compounds.
9. Application of the developed, validated method to the analysis of wastewater samples.

Objetivos

El objetivo principal de esta tesis doctoral fue investigar drogas (lícitas e ilícitas) y metabolitos en aguas residuales y superficiales utilizando técnicas analíticas avanzadas. Para alcanzar este objetivo, se han aplicado tanto la predicción del tiempo de retención, como las estrategias de “screening” de “suspect”, “non-target” y “target”. Diferentes configuraciones instrumentales han sido investigadas, incluyendo analizadores de masas con baja resolución usando espectrometría de masas en tándem (MS/MS) de triple cuadrupolo (QqQ) y alta resolución MS utilizando el cuadrupolo - tiempo de vuelo (QTOF) y la trampa de iones lineal - Orbitrap (LTQ Orbitrap).

Para alcanzar este objetivo general, se han establecido varios objetivos específicos:

- 1) Aplicación de una metodología de análisis cualitativo basada en cromatografía líquida de ultra-alta resolución (UHPLC) acoplada a QTOF MS en el modo de MS^E, lo que permite el “screening” amplio y rápido de un gran número de compuestos apoyado en una gran base de datos interna. Esta metodología se aplica a una serie de muestras de agua (influyente y efluente de aguas residuales así como aguas superficiales) de diferentes orígenes y países (Colombia, Italia y España).
- 2) El desarrollo de una aproximación sencilla para la predicción del tiempo de retención que pueda ser fácilmente aplicada durante el “suspect screening” de productos farmacéuticos y drogas ilícitas para ayudar en su identificación. Se realiza una comparación con la anchura de la ventana de extracción de masa.
- 3) El desarrollo de una segunda aproximación de predicción de tiempo de retención, más compleja y precisa, que se puede aplicar en la identificación de los metabolitos y productos de transformación para la que no están disponibles los patrones.
- 4) El desarrollo y aplicación de un flujo de trabajo combinado “suspect” y “non-target screening” mediante un software avanzado de deconvolución tras la inyección en LTQ-Orbitrap.
- 5) El desarrollo de una estrategia de análisis cuantitativo para nuevas sustancias psicotrópicas (NPS) en aguas residuales influentes basada en la extracción en fase sólida (SPE), seguida de la determinación mediante LC-MS/MS con un analizador de masas de triple cuadrupolo. También se comparan los resultados frente a un método de inyección directa. Los compuestos de interés se seleccionan en función del consumo y la prevalencia en los estudios previos toxicológicos y de epidemiología de aguas residuales.

Plan de Trabajo

Esta Tesis Doctoral se enmarca dentro de las Redes Internacionales de Formación (ITN) Marie Curie, en concreto la Red SEWPROF financiada por la Unión Europea (más información en el **Chapter 1**), lo que hizo posible la realización de diversas estancias de investigación en varios institutos internacionales.

El plan de trabajo seguido con el fin de alcanzar los objetivos señalados anteriormente es el siguiente:

Screening Cualitativo mediante Cromatografía Líquida de Alta Resolución acoplada a Espectrometría de Masas

El trabajo en esta sección se realizó en tres institutos diferentes: Universitat Jaume I, Mario Negri Institute for Pharmacological Research y KWR Watercycle Research Institute y los planes de trabajo específicos dentro de cada instituto se resumen a continuación:

Plan de trabajo en la Universitat Jaume I (España)

1. Mantenimiento continuado de la base de datos interna desarrollada por nuestro grupo para la detección cualitativa de productos farmacéuticos y drogas ilícitas. Esto implicaba incluir (al menos) el nombre y la masa molecular, junto con la masa exacta de los iones fragmento cuando se conozca.
2. Toma de muestras de aguas superficiales procedentes de diferentes zonas de la región valenciana. Esto implicó físicamente la toma de muestras en botellas de poliuretano de alta densidad, transportarlas al laboratorio y almacenarlas a 4°C.
3. Tratamiento de las muestras de aguas residuales y de aguas superficiales siguiendo la técnica de extracción en fase sólida previamente desarrollada y empleada en nuestro laboratorio.
4. Inyección de los extractos de muestra en UHPLC-QTOF en modo MS^E, lo que permite simultáneamente la recogida de datos de masa exacta de baja y alta energía.
5. Análisis de los datos mediante un software especializado, ChromaLynx (MassLynx, versión 4.1, Waters Corporation), que facilita la realización de un análisis de “wide-scope screening” mediante la extracción automática de la masa exacta de un compuesto a baja energía (y de alta energía en el caso de los iones fragmento) en forma de un cromatograma de iones extraídos (XIC).

6. Un análisis más detallado se realiza sobre los datos brutos para compuestos intactos e iones fragmento para asegurar que ningún compuesto es no detectado o identificado erróneamente.

Plan de trabajo en el Mario Negri Institute for Pharmacological Research (Italia)

1. Formación inicial teórico-práctica sobre LTQ-Orbitrap.
2. Inyección de muestras en el LTQ-Orbitrap en modo de adquisición dependiente de los datos (DDA), lo que significa que cuando un pico supera un cierto umbral preestablecido, se realizan barridos en modo MS/MS para adquirir iones producto.
3. Análisis de los datos mediante un software especializado (TraceFinder EFS, Thermo Scientific) basado en la misma base de datos utilizada en la Universitat Jaume I, donde los resultados "positivos" se basan en la masa exacta de la molécula protonada junto con el patrón isotópico, los fragmentos de iones y una búsqueda en la biblioteca (los dos últimos sólo cuando esté disponible esta información).
4. Para un examen más detallado de los compuestos "positivos", se inyectaron las muestras en el LTQ-Orbitrap en modo MS/MS, sobre la base de una lista de "target" de los compuestos "positivos" para obtener información sobre los iones producto.
5. Un análisis más profundo se lleva a cabo sobre los datos en bruto (con el QualBrowser, Thermo Scientific) para compuestos intactos y los iones producto para asegurar que ningún analito no es detectado o identificado por error.

Plan de trabajo en el KWR Watercycle Research Institute (Países Bajos)

1. Inyección de muestras en el LTQ-Orbitrap en modo de adquisición dependiente de los datos (DDA), lo que significa que cuando un pico supera un cierto umbral preestablecido, se realizan barridos MS/MS para adquirir iones producto.
2. Análisis de datos con el software especializado para tanto "suspect" (Ms Xcelerator (MsMetrix) y Sieve 2.1 (Thermo Scientific)) como "non-target" análisis (Sieve 2.1 (Thermo Scientific)).

3. Para el análisis “hidden target”, después de que se establecen los parámetros iniciales de “peak-picking”, se utiliza la misma base de datos, como en estudios previos, para detectar compuestos dentro de las muestras.
4. Además tras el “peak-picking” inicial, se utiliza la base de datos ChemSpider para detectar compuestos.
5. Todas las muestras se reinyectan en modo MS/MS en el LTQ-Orbitrap generando una lista de “target” combinando los compuestos detectados tanto en análisis “suspect” como “non-target”, para favorecer la fragmentación.
6. También se realiza un análisis de las muestras mediante QualBrowser (Thermo Scientific) para comparar los iones fragmentos observados experimentalmente con los encontrados en la literatura o en bases de datos de espectros de masas, para la identificación tentativa de los positivos.

Predicción de Tiempo de Retención

Predicción de Tiempo de Retención basado en $\log K_{ow}$

1. Revisión bibliográfica de las tendencias y los métodos de predicción de tiempo de retención actuales.
2. Desarrollo de una base de datos de tiempos de retención basada en patrones de referencia tanto de nuestro laboratorio como de la empresa Waters.
3. Obtener los códigos SMILES para todos los compuestos de la base de datos.
4. Los valores de $\log K_{ow}$ teóricos se generan mediante el software gratuito de VCLabs mientras que los valores experimentales se encuentran en la literatura y en bases de datos en línea.
5. Se realiza una regresión lineal entre los valores de $\log K_{ow}$ (tanto experimentales como teóricos) y el tiempo de retención; diferentes subconjuntos, basados en distintas familias químicas se evalúan para mejorar la correlación.
6. Investigación y comparación del impacto de un tiempo de retención predicho mediante $\log K_{ow}$ y la anchura de las ventanas de extracción sobre la eliminación de picos de falsos positivos en la detección de contaminantes emergentes (con patrones de referencia disponibles) en aguas ambientales.

Predicción de Tiempo de Retención basado en Redes Neuronales Artificiales (en colaboración con Leon Barron, King's College London)

1. Revisión bibliográfica de los métodos de predicción de tiempo de retención actuales basados en redes neuronales artificiales (ANN)
2. Los descriptores moleculares adicionales requeridos por la aproximación ANN se generan mediante el software Parameter Client (VCC Labs)
3. Optimización del método ANN basado en la aplicación de diferentes proporciones de los conjuntos de training:verification:blind, así como diferentes arquitecturas y tipos de modelo.
4. Investigación del impacto de los tiempos de retención predichos mediante ANN en la exclusión de los picos de falsos positivos en la detección de contaminantes emergentes (con estándares de referencia) en aguas ambientales.
5. Investigación de tiempos de retención predichos mediante ANN en la identificación tentativa de metabolitos y productos de la transformación para los que no se dispone de patrones de referencia.

Determinación Cuantitativa de Nuevas Sustancias Psicoactivas Seleccionadas en Muestras Influyente de Aguas Residuales

1. Selección de nuevas sustancias psicotrópicas (NPS) basada en revisión de la literatura
2. Optimización de los parámetros de MS y MS/MS a partir de una mezcla de patrones en disolvente.
 - i. Adquisición de espectros de masas en modo “full-scan” para establecer el modo de ionización y el voltaje de cono más apropiado para el ión precursor.
 - ii. Mejorar la formación de la molécula protonada mediante el uso de aditivos en la fase móvil (es decir ácido fórmico y acetato de amonio).
 - iii. Después de la optimización del voltaje de cono del ión precursor, la energía de colisión se optimiza a continuación a fin de seleccionar los iones producto más adecuados, teniendo en cuenta la sensibilidad y la especificidad (es decir, evitar transiciones no específicos tales como la pérdida de agua).
 - iv. La selección de al menos dos transiciones MS/MS por compuesto: una para la cuantificación y al menos otra para la confirmación.

3. Selección de las condiciones cromatográficas óptimas mediante la inyección de soluciones estándar, modificando la fase móvil con aditivos para mejorar la forma del pico cromatográfico y/o la separación.
4. Estudio del proceso SPE (comparación de diferentes cartuchos SPE) usando muestras blanco enriquecidas antes y después del paso por la SPE.
5. Estudiar la posibilidad de inyección directa para este método.
6. Evaluar las recuperaciones de filtración para ambos filtros de vacío y jeringa.
7. Evaluar la estabilidad de todos los compuestos a diferentes temperaturas (-20°C, 4°C y 20°C) durante un máximo de 7 días (30 días a -20°C).
8. Validación de los métodos optimizados mediante la evaluación de la linealidad, exactitud y precisión de los experimentos de recuperación a diferentes niveles de concentración. Estimación de los límites de cuantificación y detección de todos los compuestos estudiados.
9. Aplicación del método desarrollado y validado para el análisis de muestras de aguas residuales.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Emerging Contaminants in Water

Water is the most important resource on the planet, essential for all life and covers more than two-thirds of the Earth's surface; however, less than 3% of that is freshwater, able to be used for drinking. It is therefore crucial to ensure that this great minority of freshwater is kept clean, which is being achieved thanks to the introduction of wastewater and drinking water treatment plants. However, despite these measures, it is impossible to remove all unwanted compounds, resulting in them entering effluent, surface water and even in drinking water. Many of these unwanted compounds belong to the group called emerging contaminants (ECs), a wide group of compounds that are not yet regulated despite their wide presence in the environment.

The occurrence of ECs in water is a cause for concern. ECs encompass a wide variety of everyday, well-consumed anthropogenic and natural products including pharmaceuticals, drugs of abuse, pesticides, personal care products and industrial chemicals. With burgeoning populations, new chemicals being made and used combined with the impossibility of wastewater treatment plants (WWTPs) to completely remove all ECs, they can be detected in effluent wastewater, and more worryingly, surface waters. In the early 1990s, reports began marking their appearance in groundwater and drinking water and it was found that WWTPs were the major source of these ECs (Ferrer and Thurman, 2013). In the early days of environmental analytical chemistry the focus was on the contaminants emitted at higher concentrations, such as polychlorinated biphenyls (PCBs) and dichlorodiphenyltrichloroethane (DDT), due to instrumental limitations. But now, with instrumental and analytical improvements, more ECs can be detected and identified at the trace levels that they are commonly present at, albeit as a real analytical challenge.

This thesis is primarily concerned with two subsets of ECs: pharmaceuticals and illicit drugs (PIDs). PIDs still encompass a huge number of compounds, with approximately 3000 pharmaceuticals alone and almost 600 illicit drugs and new psychoactive substances (Richardson and Ternes, 2014; UNODC, 2014). In spite of these huge numbers, little is known about their individual use (or abuse), let alone their environmental impact. The big question, alluded to throughout this section, is whether the relative low concentrations of PIDs found in the environment (\sim ng/L) would cause adverse effects to either humans or wildlife. Environmental and public health organisations periodically revise their priority lists of contaminants based on health effect and occurrence in environmental waters. In the United States, the Environmental Protection Agency has recently included one antibiotic

(erythromycin) and nine natural and synthetic hormones (17α -ethinylestradiol, 17α -estradiol, 17β -estradiol, equilenin, equilin, estriol, estrone, mestranol, and norethindrone) as priority drinking water contaminants. The European Commission have included two pharmaceuticals (diclofenac and ibuprofen) and two hormones (17α -ethinylestradiol and 17β -estradiol) in their directive (Richardson and Ternes, 2014). In spite of these findings, it is important to note that it is not just these ~3600 parent compounds to investigate, but also their metabolites and transformation products (TPs), giving altogether a huge list of compounds of concern in the aquatic environment.

As a means to investigate illicit drug use, wastewater-based epidemiology (WBE), first surmised by Daughton (Daughton, 2001) and first employed four years later (Zuccato et al., 2005) has become a very interesting and continually-widening area of research.

1.1.1 Wastewater-based Epidemiology

Traditionally, surveys and interview data have been used to estimate the number of illicit drug consumers within a society. However, this comes with distinct biases and is open for conjecture. Coupled with the ever-changing world of drug use, a more real-time approach was needed, first theorised in 2001 (Daughton, 2001). WBE was first introduced in 2005 in Italy (Zuccato et al., 2005) and has been particularly utilised in Europe (Ort et al., 2014; Thomas et al., 2012) but has since expanded into China (Du et al., 2015; Khan et al., 2014), Hong Kong (Lai et al., 2013b), South Korea (Kim et al., 2015) Australia (Irvine et al., 2011; Lai et al., 2013a), Canada (Metcalf et al., 2010) and the United States (Banta-Green et al., 2009; Subedi and Kannan, 2014).

WBE initially focussed on the consumption of illicit drugs (conventionally cocaine, amphetamine, methamphetamine, cannabis and MDMA (ecstasy)) (Ort et al., 2014; Thomas et al., 2012), but has since moved, together with the drug users, to more current drugs such as new psychoactive substances (NPS) (Kinyua et al., 2015) and other lifestyle biomarkers such as caffeine, nicotine (Senta et al., 2015) and alcohol (Reid et al., 2011). Recently, further population biomarkers such as F_2 -isoprostanes (Daughton, 2012; Ryu et al., 2015) and health and exposure biomarkers including pesticides, pharmaceuticals and steroids have also been postulated (Thomas and Reid, 2011; Venhuis et al., 2014).

This approach is based on the knowledge that illicit drugs consumed by an individual may be excreted either unchanged or as one (or a number of) metabolites. From the excretion and metabolism data, the parent drug (in the case of amphetamine, methamphetamine and MDMA) or a metabolite (benzoylecgonine for cocaine and THC-COOH for cannabis) can be chosen as the biomarker. It can be assumed that if a biomarker is sufficiently stable in water and is conveyed efficiently to the WWTP,

the amount collectively excreted in a given timeframe should be reflected by the amount reaching the WWTP in the corresponding interval. (van Nuijs et al., 2011)

This approach provides objective, quantitative, near real-time profiles of illicit drug consumption by monitoring selected drug biomarkers entering the municipal sewage system. The concentration of these illicit drug biomarkers measured in influent wastewater (IWW) can be used to back-calculate the mass loads of the parent drugs and/or metabolites. These loads, together with knowledge of the drug metabolism and excretion can be used to estimate drug consumption which can then be extrapolated to the population served by the WWTP by dividing the drug consumption by the population.

While this approach sounds simple, there are a lot of underlying assumptions and limitations, including sampling procedure, characteristics of the sewer system, and the biomarker chosen (Burgard et al., 2014; Castiglioni et al., 2013). The biomarker selected should have a known excretion rate, be stable in wastewater (under a variety of weather conditions) for an extended period of time and be unique for the drug of interest. The first of these issues can be helped with toxicological information, leading to great collaborations between chemists and toxicologists within the WBE field (see Section 1.4 for more information on these collaborations). The second of these issues is one that has recently been reviewed and readers are advised to refer to McCall et al. (McCall et al., 2016). Finally, the third issue can be very troublesome. For example, amphetamine is a metabolite of methamphetamine; while MDMA and amphetamine are enantiomeric, so differentiating excreted and dumped drugs needs advanced and expensive enantiomeric profiling (Emke et al., 2014). Furthermore, in spite of heroin being one of the most concerning drugs of abuse, its analysis for WBE is very complex, as its major, stable, metabolite is morphine, which is in itself a well-known prescribed drug. However, the exclusive metabolite, 6-monoacetyl morphine has a low excretion rate, low in-sample stability and can easily transform to morphine. A reliable, unbiased biomarker for heroin has thus not yet been found (McCall et al., 2016). In spite of these limitations, WBE is a rapidly expanding field, with research spanning the world, and the traversing chemical fields: from pure analytical chemists to toxicologists, engineers, epidemiologists, statisticians, environmental chemists and computational chemists. With such a diverse range of fields and (so far) very promising and interesting results, the future is looking bright for WBE.

1.2 Analytical Strategies

WBE approaches are normally quantitative, focussing on a small number of target compounds, for which low resolution but highly sensitive LC-MS/MS approaches are perfect. However, as WBE grew, so did the knowledge that there are thousands of potential compounds of interest in any given sample, changing the focus to high resolution mass spectrometry (HRMS) instruments, capable of performing suspect screening methods based on in-house databases containing hundreds if not thousands of compounds. These instruments lead researchers beyond WBE, and into the environmental field, determining the existence of compounds in environmental matrices. Beyond suspect screening is non-target screening, which has become a significant aspect of environmental analytical chemistry, where HRMS again is a prerequisite and theorises that no previous information on potential compounds of interest is available.

1.2.1 Liquid Chromatography coupled to Mass Spectrometry

Liquid Chromatography (LC) is an affinity-based chromatographic separation technique. Simply put, it is used to separate the components of a mixture present in a sample into its individual constituents. This is based on the polarity each individual component has, and their affinity to the stationary and mobile phases. Those components that are strongly retained by the sorbent particles of the stationary phase move slowly with the flow of the liquid mobile phase, and *vice versa*. This difference in “retention time” separates the samples into its individual parts, which can then be analysed quantitatively or qualitatively.

The most common LC technique for the analysis of environmental samples is reversed phase (RP), although normal phase (NP) also has some applications as well as the intermediary hydrophilic interaction LC (HILIC). NP chromatography is used for the separation of more non-polar compounds, based on their affinity for a more polar stationary phase (i.e. silica) with a non-polar mobile phase (i.e. hexane, chloroform) used to separate those compounds more readily soluble in non-polar solvents. RP chromatography is far more common, incorporating an aqueous mobile phase and more non-polar column (based on silica stationary phases, which has been surface-modified to include long alkyl chains (e.g. C₈H₁₇, C₁₈H₃₇; called C₈ and C₁₈ columns respectively). Finally, HILIC chromatography provides improved separations and detection for highly polar compounds. The stationary phase is polar, similar to NP, incorporating silica, amino or cyano, but the mobile phase is similar to those employed in RP. HILIC also allows the analysis of charged substances, as in ion chromatography (Buszewski and Noga, 2012). In spite of its complex retention mechanisms, HILIC

is suitable for analysing compounds that elute near the void in RP, which has led some groups to use it for analysing polar contaminants in the environment. (van Nuijs et al., 2011)

The latter two chromatographic techniques also incorporate buffer solutions or additives to improve chromatographic peaks. The most common are small organic acids such as formic and acetic acid. Their widespread use is based on two fundamental reasons. First, many chromatographic separations benefit in terms of retention and/or peak shape under acidic conditions, because any silanol activity is suppressed. Second, most mass spectrometric measurements are done in positive ion mode, which is accomplished by the addition of a proton to form the protonated molecule $[M+H]^+$. Organic acids have the necessary acidity and volatility to provide an excess of protons for this purpose. This volatility is necessary because one of the functions of the LC-MS interface is to vaporise the mobile phase, meaning that these acids are preferred to the popular but non-volatile LC-UV phosphate buffers. Although organic acids are the most common mobile phase additive, it may be necessary under certain circumstances to use more neutral conditions, either because the analytes are sensitive to acids, do not exhibit optimal resolution at low pH or because ionization in negative mode is used. When acids are not suitable, volatile salts such as ammonium formate and ammonium acetate may be the additives of choice.

High performance liquid chromatography (HPLC) and ultra-high performance (or pressure) liquid chromatography (UHPLC) have both been used for the determination of emerging contaminants in water, with the latter being more highly utilised since its introduction in 2004. UHPLC has revolutionised LC, with HPLC methods able to be easily transferred. This, combined with a reduction in analysis time, has made it a powerful technique (Guillarme et al., 2010). More information relating to the use of HPLC and UHPLC in combination with high resolution mass spectrometry (HRMS) for the investigation of PIDs is found in **Scientific Article 1**.

The interface between the LC and mass spectrometer is one of the most important aspects of LC-MS. Electrospray ionisation (ESI) is almost exclusively used in the analysis of emerging contaminants in environmental waters. The electrospray interface produces singly or multiply charged ions directly from an aqueous–organic solvent system by creating a fine spray of highly charged droplets in the presence of a strong electric field, with the assistance of heat or pneumatics. The successful formation of ions using ESI requires two steps: the transfer of the compound of interest into the gas phase and the addition of a charge to the analyte if it is not already in a charged state (Taylor, 2005). This interface

can work in both positive mode ($[M+H]^+$, where the analyte gains a proton and becomes positively charged; i.e. basic compounds) and negative mode ($[M-H]^-$, where the analyte loses a proton and becomes negatively charged; i.e. acidic compounds).

The second part of an “LC-MS” instrument, mass spectrometry (MS), is the means for identification of a compound, where, after a compound has been separated chromatographically, it can be identified based on its mass. The interface converts components of the sample into rapidly moving gaseous ions, which the MS can then measure the mass-to-charge (m/z) ratio of, either in space or time, in the form of mass spectra. As most emerging contaminants are small molecules, the charge is almost exclusively 1, meaning the m/z is simply the mass of the compound, with an extra proton or less a proton, depending on whether positive or negative ionisation was used (i.e. $[M+H]^+$ or $[M-H]^-$).

The different mass spectrometers are differentiated based on their mass analysers, all of which have their own advantages and disadvantages. Some of the most common are quadrupole (Q), ion trap (IT), time-of-flight (TOF), Orbitrap and Fourier-transformed ion cyclotron resonance (FT-ICR). The main criteria used to distinguish the mass analysers are the resolution (R) (the ability to separate two peaks of similar m/z) and mass accuracy (the error associated with the measured m/z). **Table 1.1** (adapted slightly from (Krauss et al., 2010)) compares the parameters of these instruments. In order to improve sensitivity and to provide greater information than a single MS alone, the combination of two MS analysers has become prevalent, giving rise to triple quadrupole (QQQ), linear ion trap-Orbitrap (LTQ-Orbitrap) and quadrupole-time-of-flight (QTOF), all of which were utilised in the work carried out for this thesis.

Table 1.1: Comparison of commercial mass analysers in LC-MS instruments. Typical values for an m/z range of 300-400 are given; specific instruments or configurations might have improved figures

Mass Spectrometer	Resolving Power ^a (FWHM)	Mass accuracy (ppm)	Linear dynamic range	Sensitivity ^c (absolute mass)
QqQ	Unit resolution ^b	50	10^4	fg-pg
QIT	10,000	50	10^3	fg-pg
TOF	20,000	3	10^2 - 10^3	pg (full scan)
Orbitrap	100,000	2	10^3 - 10^4	fg-pg (full scan)
FT-ICR	1,000,000	≤ 1	10^4	pg (full scan)

^a: Resolving power depends on m/z range and scan speed on most instruments

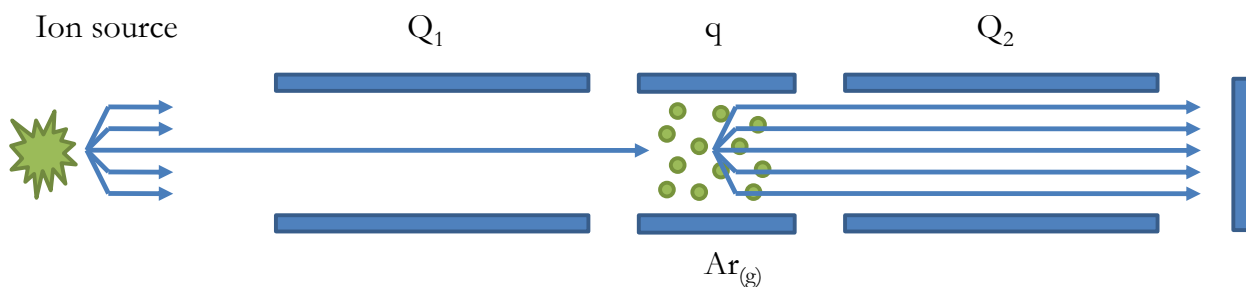
^b Unit mass resolution is the resolution for standard quadrupole instruments; with special hyperbolic quadrupole instruments a resolving power of 5,000 and a mass accuracy of 5 ppm can be achieved without a major loss of signal intensity

^c fg (femtogram, 10^{-15} g), pg (picogram, 10^{-12} g)

Sensitivity depends strongly on the ionization efficiency of the compound in the ion source

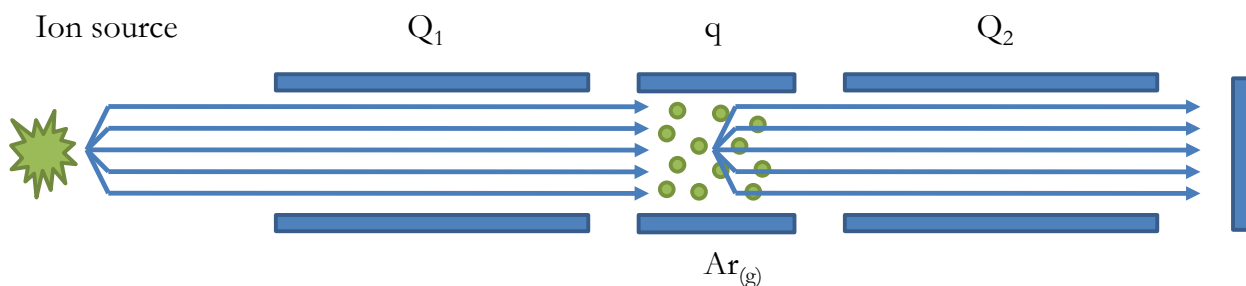
Within low resolution mass analyzers, the triple quadrupole (QqQ) is the most popular at present, and the choice for the majority of analytical environmental chemistry methods, especially because of its high sensitivity – required for the trace analysis of emerging contaminants in environmental waters – and robustness. It consists of two quadrupole mass analysers (Q) in series, with a non-mass-resolving radio frequency only quadrupole (q) in between to act as a collision cell, to aid in collision induced dissociation (CID). CID occurs based on the collision between an ion that is selected in the first quadrupole (Q_1) and an inert gas. The fragment ions produced are then passed through to the second quadrupole (Q_2) where they may be filtered or scanned. Each quadrupole can be operated independently in either full-scan or selected ion monitoring (SIM) mode. The objective of the analysis influences which mode is used by each Q, with the four most common being product ion scan, neutral loss scan, precursor ion scan and selected reaction monitoring (SRM) mode, all of which will be briefly explained below.

i) Product ion scan



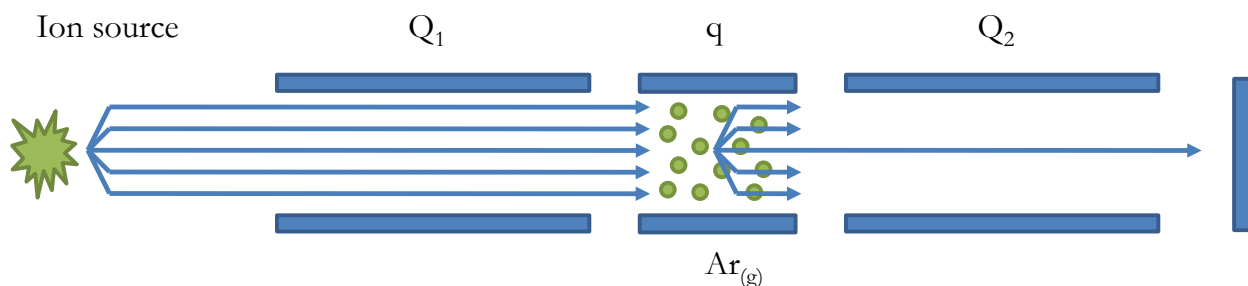
In the product ion scan, Q_1 is set to select an ion of a certain mass which is fragmented in the q . Q_2 is then set to scan the full m/z range, giving information on fragment ions of the selected precursor ion. The information obtained from this scan can be useful for the elucidation of the structure of the original ion, but is commonly used to identify the transitions used for a subsequent quantitative analysis in the method development.

ii) Neutral loss scan



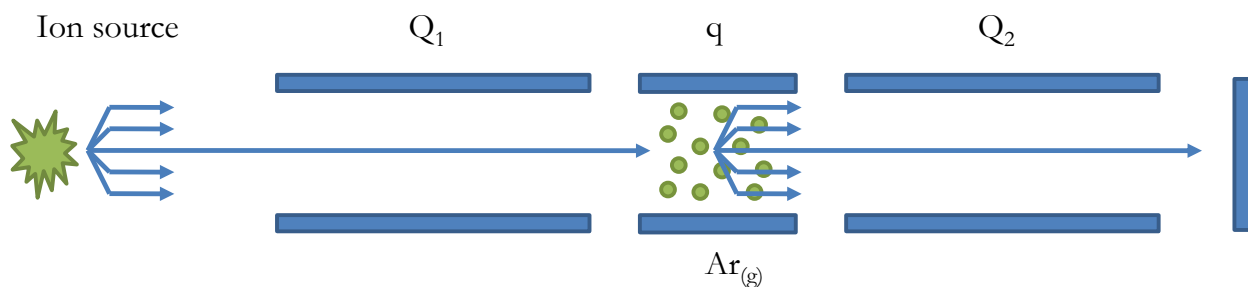
In this scan, both Q_1 and Q_2 are scanned together but Q_2 has a mass offset equal to that of a neutral loss (i.e. H_2O , CO , NH_3). This allows the selective detection of all ions which, by fragmentation in q , lead to the loss of a given neutral fragment. This technique is useful in the screening of closely related classes of compounds in a mixture or unknown samples.

iii) Precursor ion scan



When utilizing a precursor ion scan, a certain product ion is selected in Q_2 , and the precursor masses are scanned in Q_1 . This method is used for the detection of structural homologues that have a common fragment ion.

iv) Selected reaction monitoring



When employing selected reaction monitoring (SRM) mode, both Q_1 and Q_3 are set at a specific mass, allowing only a distinct fragment ion from a certain precursor ion to be detected. This is normally known as a “transition” when working in tandem MS. This working mode results in increased sensitivity, and is the approach typically used in quantitative analysis as in **Scientific Article 7**.

The use of QqQ instruments for quantitative, small scale studies does hold promise and should be a mainstay in environmental analytical chemistry laboratories. The use of these instruments in LC-MS/MS methods is nowadays the most common way to determine priority (polar) contaminants in the aquatic environment. However, with only a limited number of compounds able to be quantified in this way due to the acquisition time of each transition and the related instrumental limitations, it is insufficient for a complete overview of contaminants present in the samples.

High Resolution MS transcends these limitations and enables the detection of a theoretically limitless number of compounds, even without the need of reference standards. This is made possible with the

combination of full-spectrum data with high mass resolution and mass accuracy. More information on these instruments, combined with UHPLC is given in **Scientific Paper 1**.

1.2.2 Matrix Effects

The main limitation of LC-MS/MS quantitative strategies interfaced with ESI is matrix effects, which occur when matrix interferences coeluting with the compound(s) of interest alter the ionisation efficiency, resulting in signal enhancement or suppression. Matrix effects are the result of competition between nonvolatile matrix components and analyte ions for access to the droplet surface for transfer to the gas phase (Taylor, 2005). The magnitude of these effects are influenced by analyte and matrix properties such as proton affinity, surface activity, and polarity, as well as the ratio of analyte to matrix mass and instrument parameters including the desolvation temperature and ionization polarity of the mass spectrometer and the flow rate of the HPLC (Backe and Field, 2012).

In order to count for matrix effects, appropriate correction means are necessary, which include the use of isotope-labelled isotopic standards (ILIS), matrix-matched calibration, sample dilution, further sample clean-up and treatment and standard addition. These have all been evaluated and contain various advantages and disadvantages (Niessen et al., 2006; Stahnke et al., 2012; Taylor, 2005). It is important to take into account the specific matrix being analysed, before deciding the appropriate means of correction. For example, matrix-matched calibration or standard addition, common in pesticide residue analysis are less suitable for the analysis of drugs in wastewater. This is primarily due to the availability of a suitable “blank” sample being very difficult to obtain, let alone one that is representative of the samples.

For the analysis of illicit drugs in water samples, alternative approaches are applied to compensate for matrix effects. Most common is sample pretreatment, where solid-phase extraction (SPE) is performed to selectively retain and pre-concentrate the analytes of interest while removing interfering compounds. However, this may also lead to a pre-concentration of interfering compounds, which has led to some utilising online SPE or large volume injection (Anumol and Snyder, 2015; Farré et al., 2012; Vazquez-Roig et al., 2013). A simple approach to reduce matrix effects is sample dilution or to use smaller injection volumes. In this way, the amount of matrix introduced in the ion source is reduced, thereby decreasing signal enhancement/suppression problems generated by the interferences (Bijlsma et al., 2014; Stahnke et al., 2012).

The preferred method to compensate for matrix effects is the use of internal standards. An isotopically labelled internal standard (ILIS) is considered to be the ideal internal standard, as it shows (almost) identical behaviour to the target analyte in both sample pre-treatment, chromatography, and analyte ionization. Unfortunately, ILIS are available for only a limited number of target analytes, they are expensive, and often difficult to obtain for other target compounds, especially with sufficient isotopic purity with respect to the non-labelled variant (Niessen et al., 2006). This makes multiresidue methods troublesome, especially if some target analytes do not have an ILIS, and must therefore share with an ILIS of another compound that has a similar structure or retention time. In spite of the assurances of ILIS, even with their availability, in complex matrices it is generally recommended to also incorporate an additional means of matrix effect reduction, such as a pretreatment step in order to gain maximum sensitivity and precision.

1.2.3 Target, Suspect and Non-Target Screening

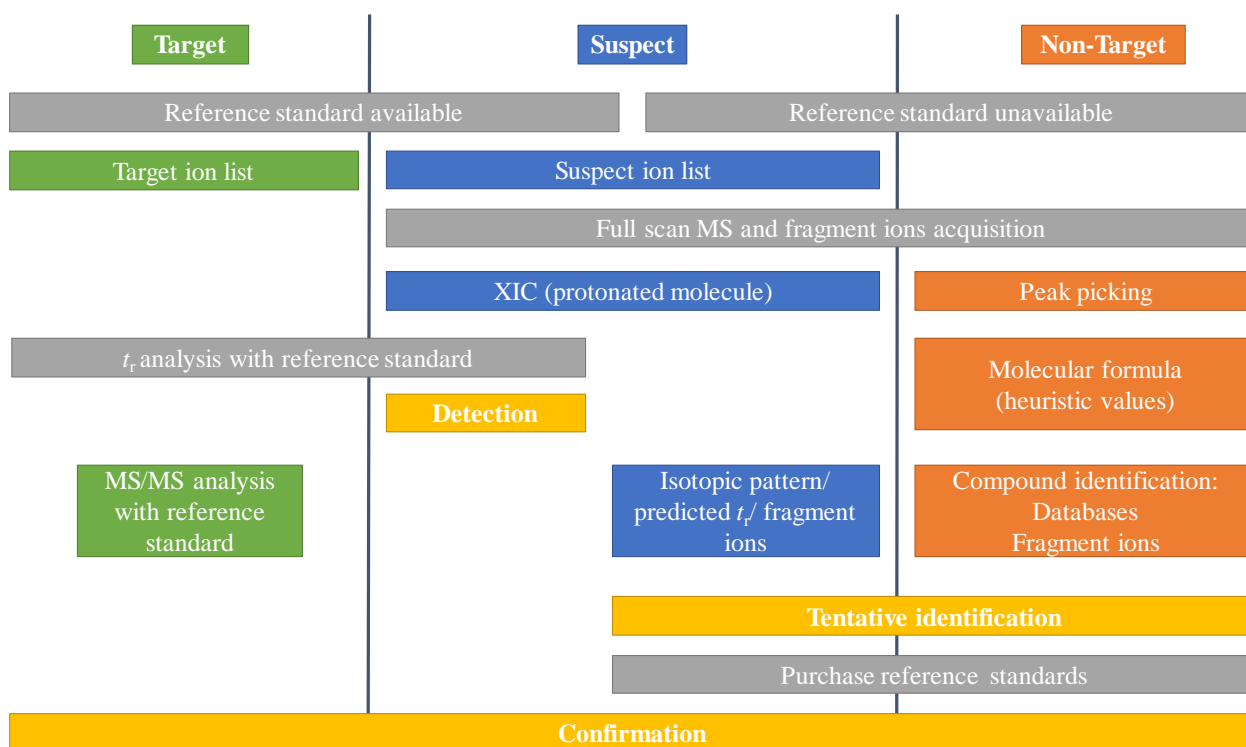


Figure 1.1: Workflow for target, suspect and non-target screenings. Each method has its own individual colour: target (green), suspect (blue) and non-target (orange), with each step specific for any particular method coloured respectively. Steps concerning more than one method are in grey, while levels of confidence of identity obtained are in yellow.

Target screening is the common approach when using the SRM mode of LC-MS/MS QqQ instruments, described above. In this method, a limited number of compounds are selected in advance, their specific transitions monitored and their presence in samples quantified based on their respective reference standards. The big limitation of this method is that there are many other compounds potentially present in samples being ignored. Some authors have proposed the term pre-target analysis for the methods based on QqQ, where the analytes must be selected before analysis, and therefore data acquisition is analyte-specific (Hernández et al., 2005).

The limitation in the number of compounds to be investigated by QqQ, has led to a shift from the traditional LC-MS/MS pre-targeted approach to HRMS (QTOF and Orbitrap) suspect and non-target approaches. These two methods rely on the high resolution, exact mass capabilities of these instruments, as no reference standards are needed, at least in the initial stages. As well as HRMS instruments, these methods also require powerful deconvolution software.

As in the famous Donald Rumsfeld speech, these three screening methods can be thought of as analysing known knowns (target), known unknowns (suspect) and unknown unknowns (non-target). In other words, target methods are specific, analysing compounds that are being deliberately searched for; suspect methods search for a much larger range of compounds (i.e. through the use of a database) but it is not known which of these compounds will be encountered in the samples; finally in non-target methods it is not known which compounds are in the samples or are being searched for. The steps associated with each method are shown in **Figure 1.1**.

Suspect analyses investigate a list of compounds that are potentially present in samples on the basis of a database, containing name, molecular formula (or exact mass), as a minimum requirement. Further information, such as fragment ions and adduct formation can also be added. The main difference with pre-target analysis is that availability of reference standards is not a prerequisite, as the information acquired is the accurate-mass full-spectrum, which implies that HRMS is needed. In this way, the selection of analytes to be searched in the samples can be made after mass data acquisition, a reason why some authors have named it as a post-target screening (Hernández et al., 2005). The data acquisition is a very important aspect of suspect analyses, with simultaneous low- and high-energy acquisition highly recommended. This HRMS acquisition is known as MS^E, all-ions MS/MS or all-ions fragmentation, depending on the manufacturer, and means that in a single run, data can be acquired relating to the parent compound (low energy) and fragment ions (high energy).

After data acquisition by HRMS, specialised, manufacturer-specific software such as ChromaLynx (Waters), TraceFinder (Thermo Scientific), DataAnalysis (Bruker) and MassHunter PCDL Manager (Agilent Technologies) incorporates the database and extracts all “positive” findings based on the user-set parameters in the form of extracted ion chromatograms (XICs). In the cases where reference standards are available, the suspect compounds can be identified based on a comparison of the parent ion (at exact mass), fragment ions (for confirmation, where possible) and retention time.

As reference standards are not necessarily available for all compounds, further steps can be used to gain tentative identification. Initially, the mass accuracy of the HRMS instruments can be utilised, with accuracy possible down to <5ppm giving great confidence to (at least) the molecular formula of a compound within the database. However, there is a distinct possibility of isobaric compounds (e.g. with the same exact mass), depending on the width of the extraction window used within the software, and even at a certain m/z there is a chance of an isomeric compound within the XIC. Isotopic pattern

fitness or retention time prediction can then be used to gain further identification confidence. Alternatively, fragment ions can be searched and compared with literature or online mass spectral databases. Even though reference standards may be unavailable, a high degree of confidence in the “tentative” identification can be found, based on fragment ions and their comparability and compatibility with literature and possible chemical structure, respectively.

Some groups have introduced confidence levels for the detection and identification of compounds (Schymanski et al., 2014). However, for the sake of clarity and continuity within this thesis, the criteria set out in **Scientific Article 2** are followed:

Detection: based on the presence of one accurate mass ion and retention time agreement (with reference standard)

Confirmation of the identity (shortened to confirmation), with at least two accurate mass ions and retention time agreement (with reference standard)

Tentative identification: with at least two accurate mass ions justified by literature data and/or compatibility with candidate chemical structure

In all cases, the isotope pattern observed must fit the theoretical expected from the presence in the molecule of relevant atoms such as carbon, chlorine, bromine, among others. The parameters for mass error may differ depending on instrument used, with Orbitrap tending to have better mass accuracy and therefore the acceptable mass error may be smaller than that of a QTOF. However, within QTOF instruments, depending on the vendor as well as the age and use of the instrument, maximum allowed mass errors may differ from 20ppm (Kinyua et al., 2015) down to 5ppm, the latter being used in all works performed in this Thesis (**Scientific Articles 2 and 3**).

With the ever-improving manufacturer-specific software, it is becoming more feasible to perform non-target analyses (such as the identification of metabolites or TPs) on a single platform, such as Metabolynx (Waters), Sieve (Thermo Scientific), Metabolite Tools (Bruker) and Metabolite ID (Agilent). However, it is impossible to include all necessary parameters in a single software, with complementary programs commonly sought.

Non-target analyses begin with a simple hypothesis: “I do not know anything about this sample and the possible contaminants therein, but I want to know which compounds are there.” While this does indeed sound simple and achievable, in practice it is very complex. Unlike the previous two methods,

where some sort of *a priori* knowledge is used in the development of the “target” or “suspect” lists, in non-target analysis, there is no previous knowledge at all, as implied by the name. To begin the analysis procedure, the chromatographic peaks need to be picked (i.e. which peaks are the most important). This determination can be made based on the end result: if the analyst wants to look for compounds of a high concentration, a high threshold will be given for the peaks requiring further analysis, thereby removing all peaks under the set threshold. Alternatively, if the analyst is searching for compounds, TPs or metabolites of a likely low concentration, a low threshold will be set. Nevertheless, for both methods, a subtraction based on a (procedural) blank sample is made to remove peaks known to not belong to the sample.

Once the peaks for further investigation are found, molecular formulae are given, based on various heuristic filters (Kind and Fiehn, 2007). These can then be searched in databases such as ChemSpider, PubChem, NIST and DAIOS to give an idea of the possible compound of interest (Bletsou et al., 2015; Little et al., 2012; Müller et al., 2011). When searching databases based on a molecular formulae, there is a high likelihood of isomers and therefore a ranking system is necessary to determine which the compound of interest is. This can be made based, simply, off the number of available sources (with the higher number of sources indicating a more common/known compound), or by utilising the HRMS data originally obtained, i.e. the fragment ions generated. A very useful non-target tool has been developed for this purpose, MetFusion. This software takes the fragment ions obtained for a specific peak and those computationally derived from its sister software MetFrag, compares them with the online mass spectra database, MassBank, and gives a ranked list of possible candidates based on spectral similarity (Gerlich and Neumann, 2013). The candidates given, based on either the ranking system or through the use of fragment ions can then be compared with literature data to give tentative identification or lead to the purchase or synthesis of standards can lead to confirmation.

While the methodology seems straightforward, there are distinct difficulties associated with non-target analyses. The most complex problem is that the compounds traditionally searched for using this approach are those of lower concentration and not necessarily known or incorporated into suspect databases. The peaks would therefore be in the “noise” range of most chromatograms, so powerful software is required to help deconvolute the chromatograms. Furthermore, even after distinct masses are found, there is a very high probability that there is more than one compound relating to any one specific m/z , and even isomers. For these reasons, non-target approaches are a real analytical challenge, very time-consuming and may only yield limited results. Nevertheless, this approach is gaining in

scientific interest, and with advances in the sensitivity of high resolution instruments combined with even more powerful software, they could become a routine part of environmental analytical chemistry.

1.3 Scientific Article 1

**INVESTIGATION OF PHARMACEUTICALS AND
ILLICIT DRUGS IN WATERS BY LIQUID
CHROMATOGRAPHY-HIGH-RESOLUTION MASS
SPECTROMETRY**

F. Hernández *, M. Ibáñez, R. Bade, L. Bijlsma, J.V. Sancho

Trends in Analytical Chemistry (2014) 63: 140-157

Trends in Analytical Chemistry 63 (2014) 140–157



Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry



F. Hernández*, M. Ibáñez, R. Bade, L. Bijlsma, J.V. Sancho

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

Abstract

Mass spectrometry (MS) coupled to chromatography has been increasingly used in the environmental field, as it allows the best performance currently attainable for the investigation of a wide range of organic pollutants. When dealing with emerging contaminants, there is a clear trend towards the use of liquid chromatography-mass spectrometry (LC-MS) techniques, from tandem (low-resolution) MS to high-resolution (HR) MS. HRMS allows targeted and untargeted analysis, thanks to full-spectrum acquisition at high mass accuracy with good sensitivity. With the same instrument, target, suspect and non-target screening can be performed, as can retrospective analysis and discovery of transformation/degradation products. This article gives a general overview on the use of HRMS in LC-based methods directed towards the investigation of illicit drugs and pharmaceuticals in the aqueous environment. We consider time-of-flight and Orbitrap mass analyzers, and we discuss the benefits of using ultra-high-performance liquid chromatography (UHPLC) in combination with HRMS.

Keywords

Emerging contaminant; High-resolution mass spectrometry (HRMS); Illicit drug; LC-HRMS; Liquid chromatography (LC); Orbitrap; Pharmaceutical; Screening; Time of flight; Transformation product

1. Introduction

The increasing number of emerging contaminants, such as pharmaceuticals (both human and veterinary) and personal-care products (PPCPs), or illicit drugs, detected in the water cycle can be attributed to the growth of the human population, the shift towards the use of more hydrophilic compounds in consumer applications, and undoubtedly the improvements in selectivity and sensitivity of modern analytical techniques. Pharmaceuticals and illicit drugs (PIDs) are continuously excreted or discarded into the sewer systems as the unaltered parent compounds or their metabolites. Subsequently, they often end up in environmental waters, as a consequence of incomplete elimination by wastewater-treatment plants (WWTPs) [1]. There is justified concern over the possible impact of these pharmacologically-active compounds on the environment, especially over the long-term toxicological effects on living organisms and the combined effect of exposure to multiple compounds, particularly antibiotics [2]. In addition, a large number of transformation products (TPs), in many cases unknown, can be formed in the water cycle and need to be taken into account to know the overall contribution of these contaminants in the environment [1]. The presence of PIDs and their metabolites in water and sediments has spurred researchers to set up monitoring studies to evaluate their fate [2] and [3], and their removal and transformation in WWTPs [4], [5] and [6]. An interesting field that has emerged from the analysis of wastewater is the so-called sewage-based epidemiology (SBE), directed towards the estimation of illicit drug use of a population on the basis of the determination of appropriate biomarkers in influent wastewater. SBE is a promising, complementary tool to existing population surveys and other conventional approaches to estimating illicit drug use. SBE requires the application of sophisticated analytical methodologies able to quantify illicit drugs accurately in urban raw wastewater [7], [8] and [9].

The very low concentrations generally found for PIDs and/or their metabolites (commonly sub- $\mu\text{g/L}$ levels), in combination with the complexity and the unknown composition of the different aqueous matrices to be analyzed, makes it necessary to use highly sensitive, selective analytical methodologies. The majority of the methods developed have been based on liquid chromatography (LC) coupled to tandem mass spectrometry (LC-MS/MS), particularly with triple quadrupole (QqQ) analyzers [2] and [10]. The medium-high polarity of most PIDs justifies the predominance of this technique. The use of MS/MS under selected-reaction monitoring (SRM) mode facilitates the accurate quantification of target analytes at trace levels. Furthermore, the acquisition of two SRM transitions, together with retention-time data and measurement of ion-intensity ratios, gives sufficient information for safe identification. Typically, the most sensitive transition is selected for quantification and at least one additional transition is acquired to render a reliable confirmatory method. However, the MS source and the conditions of the analyzer need to be optimized for determination, carefully considering the specificity of transitions to avoid potential false negatives or positives [11].

The current trend in analytical chemistry is the development of LC-MS based methods that seek the simultaneous determination of many compounds in a single run, providing considerable information about their occurrence, and reducing analysis time and cost [10], [12], [13] and [14]. In spite of the high sensitivity and selectivity reached, LC-MS/MS has some limitations regarding multi-class analysis, especially when broadening the scope of the method to a large number of compounds. In MS/MS methods, the acquisition time of each transition restricts the number of target analytes that can be monitored. Although latest-generation QqQ instruments have low dwell times and allow notable increase in the number of transitions acquired within a run, there is still a limitation when dealing with thousands of contaminants that may potentially be present in waters. Thus, the application of target LC-MS/MS methods is clearly insufficient to achieve a realistic, extensive overview of pollutants present in the samples. This is because, in target LC-MS/MS methods, compounds other than the selected analytes are commonly ignored, even if they are present at high levels in the sample.

High-resolution MS (HRMS) transcends this major limitation of targeted MS/MS analysis. HRMS instruments [e.g., time-of-flight (TOF) and Orbitrap] provide high-quality information by combining sensitive full-spectrum data with high mass resolution and mass accuracy [15] and [16]. In theory, the presence of an unlimited number of compounds can be investigated, without requiring the pre-selection of analytes or even without having reference standards available.

With the ever-improving technology of LC-MS systems, it is easy to overlook the equally important aspect of chromatography. However, an efficient chromatographic separation is essential to avoid or to minimize matrix interferences, and to get reliable identifications. Ultra-high-performance (pressure) LC (UHPLC) has emerged as an innovative, powerful separation technique based on the use of columns containing stationary-phase packing with particle sizes smaller than those of conventional HPLC [17]. It has led to evident improvements in chromatographic separations, an important and relevant aspect in LC-MS analysis. Shorter chromatographic run times and improved sensitivity are common advantages derived from the use of UHPLC, but some other aspects may also be considered, as highlighted in this review. This article discusses the application of LC-HRMS for the investigation of PIDs and metabolites/TPs in water samples, and the role that UHPLC can play within this field of analysis.

2. High-resolution mass spectrometry

HRMS circumvents the main drawback of targeted SRM analysis (i.e., missing non-pre-selected compounds even if they are present at high concentrations in the samples). HRMS instruments (e.g., TOF and Orbitrap) provide high-quality data by combining sensitive full-spectrum mass data with high mass resolution and mass accuracy. Since acquisition is not targeted, any ionizable compound in the sample can be, in principle, detected and investigated. Furthermore, hybrid instruments, such as quadrupole-TOF (QTOF) and Linear Trap Quadrupole Orbitrap (LTQ-Orbitrap), offer additional information on compound confirmation and/or structure elucidation [18].

Several studies employ a post-targeted approach, which involves an initial data inspection, based on the use of the exact mass of known substances in a customized database. Hundreds of compounds can be investigated on the basis of their theoretical exact mass, compared with the accurate-mass measurements, which greatly increase the reliability of identification [19]. Moreover, the presence of compounds initially not considered, such as new substances and TPs, can be investigated from full-scan accurate-mass data acquired at any time without the need for additional analysis [20], [21], [22] and [23]. This ability is advantageous, as, in many occasions, the samples may already have been discarded or the analytes degraded, so additional sample injections may not be possible. This retrospective analysis enables the screening to be further widened, by reprocessing only raw data. The main benefit of using HRMS is that it eschews the need for reference standards, as tentative identifications of suspect compounds can be made on the basis of the information provided by this technique. Obviously, reference standards are required for ultimate confirmation, but they may be acquired in a final stage when solid well-founded evidence exists on the presence of the compound in the sample. In this way, laboratories do not need to acquire all reference standards before analysis, with the subsequent problems of availability (e.g., TPs), costs and expiry dates [24].

Non-target analysis may also be explored when using HRMS. A true “unbiased” non-target screening, without any *a priori* information on the compounds to be detected, is an analytical challenge, as the process needs expertise, and is complex and time-consuming [15], [25] and [26]. An intermediate situation between target and true non-target analysis is the application of “biased” non-target approaches, where, for example, the formation of “unknown” TPs from a given parent compound is computationally predicted [27] and [28], or tentatively identified by performing laboratory experiments [22], [29], [30], [31] and [32]. Other options (e.g., searching for common fragments or mass-defect filtering) are also feasible and useful in investigating the presence of related compounds. Here, the number of chemically meaningful structures, which can be assigned to an unknown peak, is limited to structures showing a close relationship with the parent compound [15].

Krauss et al. gave an overview of the state of the art and future trends of LC-HRMS applied to the environmental analysis of polar micropollutants [15], showing illustrative systematic workflows for different approaches: quantitative target analysis with reference standards; suspect screening without reference standards; and, non-target screening of unknowns (Fig. 1).

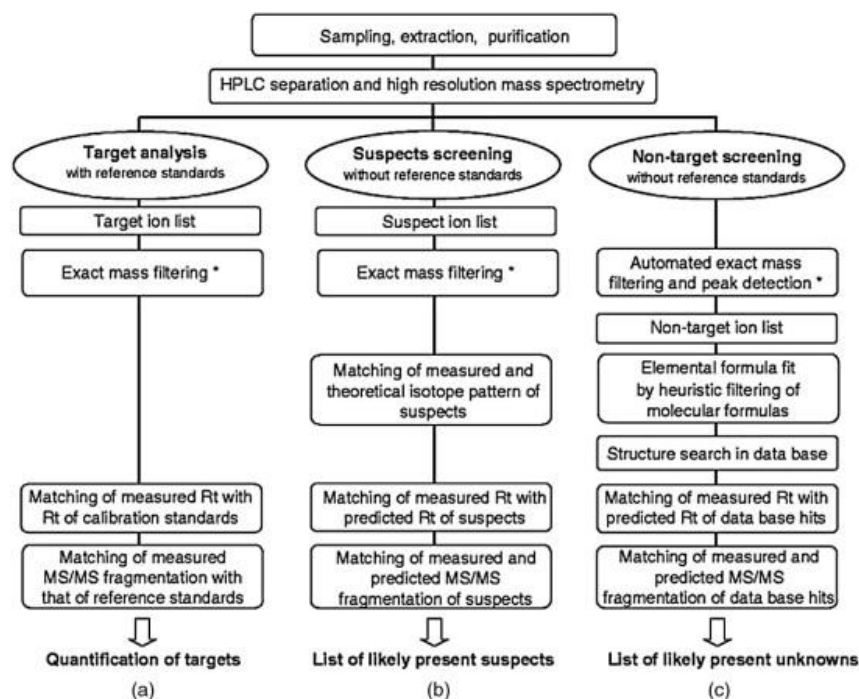


Fig. 1: Comparison of systematic workflows for (a) quantitative target analysis with reference standards, (b) suspects screening without reference standards, and (c) non-target screening of unknowns in environmental samples by using liquid chromatography-high resolution (tandem) mass spectrometry [LC-HRMS(MS)]. *Note that the m/z range of the extraction window for the exact-mass filtering depends on the mass accuracy and the resolving power of the mass spectrometer used. {Reproduced from [15] with permission}.

After a compound (based on a suspect or non-target approach) is discovered, the next step is to confirm its identity. Usually, this process involves acquisition of full product-ion spectra after re-analyzing the sample by MS/MS systems (i.e. hybrid QTOF or LTQ-Orbitrap), in order to match the observed accurate-mass product ions with the chemical structure of the suspect or candidate(s). In order to obtain fragmentation information in a single run, some hybrid analyzers allow acquisition of full-scan spectra with and without applying collision energy in a sequential fashion. Using this acquisition mode, named MS^E by Waters, or high-collision dissociation (HCD) by Thermo [33] and [34], two separate acquisition functions are sequentially measured in full-scan mode. The first, without applying collision energy in the cell (LE), obtains a conventional full spectrum, where intact (de)protonated molecules/adducts are commonly observed, followed by a second (HE), where a fixed or collision-energy ramp is applied in order to induce ion fragmentation. In this way, fragmentation information is obtained in advance for all compounds in a single run without the need to re-inject the sample in MS/MS mode.

Despite its qualitative potential, HRMS typically shows lower sensitivity than QqQ instruments operating in SRM mode, and quantitative LC-HRMS applications are more limited. However, Orbitrap and the latest TOF instruments show improved sensitivity and wider linear dynamic range, similar to that of QqQ, so prompting their use for quantification and identification/confirmation in a single run [20], [35] and [36].

3. Contributions of UHPLC

UHPLC is the most important development in LC in recent years. This technique uses chromatographic columns packed with particles with diameters below 2 μm , which significantly increase efficiency, even at high mobile-phase flow rate, achieving faster separations [17] and [37]. Conventional HPLC runs can be relatively long, particularly to avoid or to minimize co-elution of matrix interferences that may lead to difficulties in terms of identification. To this end, UHPLC can be of additional value. First, it shortens the run time, and, second, it facilitates attaining sufficient chromatographic resolution to minimize co-elution of compounds with close m/z values. In addition, a remarkable increase in detectability can be achieved as a consequence of the narrower, higher peaks provided by this technique.

The combination of UHPLC with MS appears to be a suitable approach that fulfils sensitivity, selectivity and peak-assignment certainty. However, due to the very narrow peaks produced by UHPLC (commonly 1–6 s), coupling with MS devices may be critical. For this reason, specific quadrupole-based instruments that show improved acquisition rates were launched for UHPLC coupling. The short dwell times (low-ms range) offered by modern QqQ instruments allow easy coupling to UHPLC due to the narrow chromatographic peaks obtained.

Initially, UHPLC found many applications in pesticide residues, where a large number of pesticides could easily be separated and measured by MS/MS in less than 10 min [38], [39] and [40]. Obviously, UHPLC-MS/MS has also been applied in other fields, including the quantitative determination of PIDs in water [14], [41], [42] and [43].

Aside from triple quadrupole (QqQ) mass analyzers, TOF instruments afford fast full-spectral acquisition rates at good sensitivity and high mass-resolution (10,000–40,000 FWHM depending on the instrument) with high mass accuracy (typically lower than 5 ppm). The scan speed attainable by recent accurate-mass TOF analyzers falls in the range of 10–100 scans/s, more than enough to follow the narrow chromatographic peaks obtained under UHPLC separations, without compromising mass resolution and sensitivity.

Modern TOF analyzers represent a promising alternative, particularly for screening purposes, to the well-established QqQ instruments, as TOF can deal with the large number of compounds to be searched in environmental research. The improved chromatographic resolution and detectability achieved by UHPLC separations were particularly essential when using QTOF mass analyzers in MS^E acquisition mode. As fragmentation is promoted in HE function, without pre-selecting any precursor ion, recognizing which ions are fragment ions and which are not becomes essential to avoid spectral interferences that would complicate the identification process. With this aim, UHPLC has become a valuable tool for choosing ions obtained from the HE function that perfectly co-elute with the intact (de)protonated molecule from the LE function, and

assigning them as potential fragment ions. In this way, an additional MS/MS acquisition can be avoided, speeding up the screening and confirmation steps following the MS^E approach.

In order to illustrate the benefits of UHPLC separations along this process, Fig. 2a shows the LE and HE mass spectra for a wastewater-sample extract obtained after solid-phase extraction (SPE). The LE mass spectrum (bottom) shows several abundant ions, which are expected to be the “precursor ions” of the fragments, observed in the HE mass spectrum (top). In order to correlate the potential (de)protonated molecules with their fragment ions, extracted ion chromatograms (XICs) were obtained for all ions (at LE and HE) and chromatographic peak shapes and retention times were evaluated. As can be seen in Fig. 2b, two co-eluting compounds appeared in this sample, but, based on the improved chromatographic profile achieved by UHPLC, their fragment ions could be easily differentiated. Thus, the fragment ions observed in HE with m/z 119 and m/z 91 were assigned to the “precursor” m/z 192, while the ions m/z 207 and m/z 159 were assigned to m/z 297.

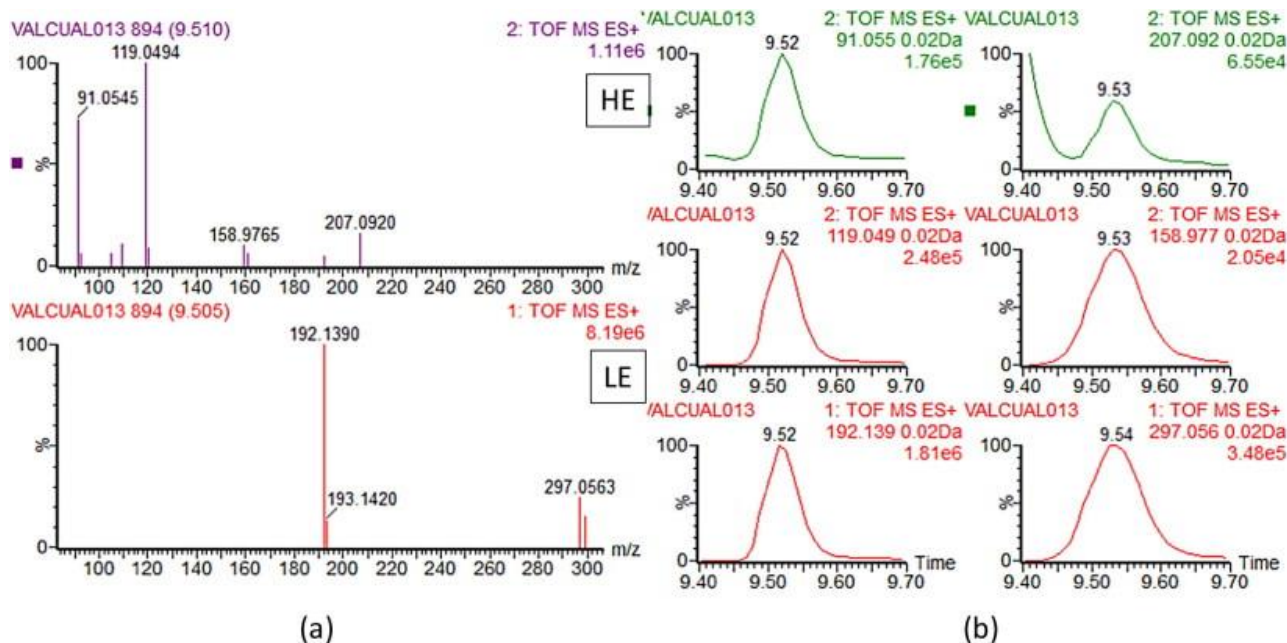


Fig. 2: (a) Low-energy (LE) and high-energy (HE) (Q)TOF mass spectra for a wastewater-sample extract after SPE; (b) XICs for the protonated molecule (at LE) and main fragment ions (at HE) for two co-eluting compounds.

A new type of HRMS analyzer, Orbitrap, was invented by Alexander Makarov in 2003 [44] and [45]. This device shows high mass resolution ($>100,000$ FWHM), high mass accuracy (<5 ppm) and acceptable dynamic range (10^3). The main drawback is its scanning speed, which is inverse of mass resolution. For example, only 1 scan/s can be acquired when using a resolution of 100,000. This limitation obviously affects the number of points per chromatographic peak and therefore its correct chromatographic peak shape when coupled to UHPLC, where peak widths are only a few seconds. By contrast, when a faster scanning speed is selected (e.g., 10 scans/s), resolution decreases dramatically (e.g., 10,000 FWHM). Thus, a compromise between achievable resolution and adequate chromatography must be found [46] and [47]. Nevertheless, several examples of the use of UHPLC columns combined with Orbitrap mass analyzers can be found [48], [49], [50] and [51]. Unfortunately, the mass resolution used is not reported or the chromatographic peak widths estimated from the chromatograms shown are unacceptably large (20–60 s) for a genuine UHPLC separation. Apparently, only Pinhancos et al. [48] worked in pseudo-UHPLC conditions, reporting chromatographic peaks of around 10s width.

In summary, so far, insufficient data have been reported to support the efficient combination of UHPLC with Orbitrap in this field. We expect further improvements in the scan speed of these mass analyzers without significantly affecting the mass-resolving power in the near future, making possible successful coupling to true UHPLC separations.

4. Applications of LC-HRMS

In this section, we present different approaches to the investigation of PIDs in waters by LC-HRMS, including target, suspect and non-target screening, and characterization of TPs of PIDs subjected to degradation under laboratory-controlled conditions.

Table 1 and Table 2 give overviews of the literature on LC-(Q)TOF-MS and LC-(LTQ)Orbitrap-MS applications, respectively.

Table 1: Literature overview on LC-QTOF-MS applications

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
Marchese (2003)	HPLC	Pharmaceuticals (5)	T	SW	Quantification and confirmation (QqQ <i>versus</i> QTOF)	[52]
Stolker (2004a)	HPLC	Pharmaceuticals (13)	T	SW, GW, DW	Quantification and confirmation (QqQ) and confirmation (QqQ <i>versus</i> QTOF)	[56]
Stolker (2004b)	HPLC	Pharmaceuticals (5)	T	SW, WW	Quantification and confirmation (QqQ <i>versus</i> QTOF)	[57]
Agüera (2005)	HPLC ^d	Diclofenac	D	A	Elucidation of Hy and PhN TPs by LC-TOF MS+GC-MS (13)	[97]
Boreen (2005)	Infusion ^d	Pharmaceuticals (5)	D	SW	Elucidation of PhN TPs of sulfa drugs by LC-UV + TOF-MS + FTIR + ¹ H NMR + ¹³ C NMR (5)	[88]
Eichhorn (2005)	HPLC	Trimethoprim	D	WW	Elucidation of Bd TPs (2) by LC-QTOF-MS + LC-IT-MS + H/D exchange. Quantification by LC-MS	[89]
Ibáñez (2005)	HPLC	Unknown compounds	NT	SW, GW, W	Screening and elucidation of unknown compounds	[25]
Petrovic (2006)	UHPLC	Pharmaceuticals (29)	T	SW, WW	Screening, quantification and confirmation	[71]
Pozo (2006)	HPLC	Antibiotics (16)	T	SW, GW	Quantification and confirmation (QqQ) and confirmation (QqQ <i>versus</i> QTOF)	[11]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D			Ref.
			^a	Matrix ^b	Observations ^c	
Martínez Bueno (2007)	HPLC ^d	Pharmaceuticals and metabolites (48), and other emerging contaminants (56T, 5S)	T, S, NT	WW	Quantification (QTRAP) and confirmation (QTRAP <i>versus</i> TOF)	[64]
Pérez (2007)	UHPLC	Enalapril and its metabolite enalaprilat	D	A	Elucidation of PhS TPs (3 from enalapril + 1 from enalaprilat by QTOF+QqLIT MS)	[98]
Farré (2008)	UHPLC	Pharmaceuticals (28) + phytoestrog ens (4)	T	SW, WW	Quantification and confirmation (QqQ <i>versus</i> QTOF)	[55]
Ibáñez (2008)	UHPLC ^d	Organic pollutants	NT	SW, WW	Screening and identification. Empirical library (104) <i>versus</i> Theoretica l library (500)	[114]
Kosjek (2008)	UHPLC	Diclofenac	D	A, HPLC	Elucidation of Bd TPs (4). Isotopic cluster analysis	[87]
Radjenović (2008)	UHPLC	Atenolol and glibenclamide	D	WW	Elucidation of Bd TPs (1 from atenolol + 1 from glibenclamide). 1 TP detected in WW by QqLIT-MS.	[107]
Ibáñez (2009)	UHPLC	Pharmaceuticals (62)	T	SW, WW	Screening and confirmation	[69]
Lavén (2009)	HPLC	Pharmaceuticals (15)	T	WW	Screening, quantification, confirmation and removal efficiency	[72]
Ferrer (2010)	HPLC	Chlorine- containing compounds	Semi-NT	DW, GW, WW, SW, HPLC	Lamotrigine and its metabolite (2-N- glucuronide) detected	[112]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
					in DW, SW and WW samples. Quantification	
Gómez (2010)	HPLC	Pharmaceuticals, transformation products and pesticides	NT	WW, SW	Screening, quantification and confirmation. Database containing pharmaceuticals (87) + pesticides (300)	[74]
López-Roldán (2010)	UHPLC ^d	Pharmaceuticals (28) and estrogens (10)	T	SW	Quantification (QqQ) and confirmation (TOF)	[65]
Magnér (2010)	UHPLC	Pharmaceuticals (10)	T	SW, MW	Screening, quantification and confirmation	[67]
García-Galán (2011)	UHPLC	Sulfamethazine	D	A	Elucidation of TPs after Bd by fungus <i>Trametes versicolor</i> (4). Removal efficiency in SS by QqLIT-MS	[108]
Gómez-Ramos (2011a)	HPLC	Sulfamethoxazole	D	n.d.	Elucidation of Oz TPs (6)	[105]
Gómez-Ramos (2011b)	HPLC	Pharmaceuticals, illicit drugs, TPs and other organic pollutants	NT	WW	Screening and identification. Database containing 147 compounds. Identification of related and unexpected TPs	[115]
Hernández (2011a)	UHPLC	Pharmaceutical metabolites (160)	S	WW	Screening and identification (retrospective analysis)	[21]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D			Ref.
			^a	Matrix ^b	Observations ^c	
Hernández (2011b)	UHPLC	Illicit drugs and metabolites (11 T + 65 S)	T, S	WW	Screening and identification	[73]
Nurmi (2011)	UHPLC ^d	Pharmaceuticals (16) and pesticides (68)	T	WW	Screening, quantification and confirmation	[120]
Pérez-Parada (2011)	HPLC	Amoxicillin	D	WW, SW	Elucidation of Hy TPs (4). 1 TP detected in SW samples	[91]
Postigo (2011a)	UHPLC	Methadone	D	A, HPLC	Elucidation of Hy, PhN and Pc TPs (6). 1 confirmed with reference standard	[100]
Postigo (2011b)	UHPLC	Cocaine	D	A, HPLC	Elucidation of Hy, PhN and Pc TPs (14)	[101]
Terzic (2011)	UHPLC	Polar contaminants	NT	Sed	The aquatic sediment was influenced by pharmaceutical industry	[111]
Díaz (2012)	UHPLC	Pharmaceuticals, illicit drugs, TPs and other organic pollutants (231 T + 1100 S)	T, S, NT	WW	Screening and identification	[16]
Eichhorn (2012)	UHPLC	Sildenafil (Viagra) and its N-demethylated metabolite	D	HPLC,A,SW	Elucidation of PhS TPs (8 from sildenafil + 6 from N-demethylsildenafil) by LC-ESI-TOF-MS + LC-APCI-QqQ-MS + H/D exchange	[90]
Ferrer (2012)	HPLC	Pharmaceuticals and degradation products (100)	T	DW, GW, SW WW	Screening, quantification and confirmation	[36]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
Gonzalez- Mariño (2012)	HPLC	Illicit drugs and metabolites (24 T + 130 S)	T, S	WW	Screening, quantification and identification	[35]
Leknes (2012)	UHPLC ^d	Oseltamivir and its metabolite oseltamivir carboxylate	T	WW	Screening, quantification and confirmation	[66]
Martínez Bueno (2012)	HPLC	Pharmaceuticals and illicit drugs (10 T, 1200 S)	T, S, NT	SW	Screening, quantification and identification	[116]
Nurmi (2012)		Pharmaceuticals and pesticides (88 T + 201 S)	T, S, NT	WW	Evaluation of T, S, NT screening with spiked samples. Application of T, S to real WW. Theoretical library (6)	[117]
Bijlsma (2013)	UHPLC	Cocaine and its metabolite benzoylecgonine	D	SW	Elucidation of Hy, Cl and PhS TPs (16 from cocaine + 10 from benzoylecgonine). 7 TPs detected in SW and WW samples by UHPLC-QqQ-MS	[30]
Boix (2013)	UHPLC	Omeprazole	D	SW	Elucidation of TPs after Hy, Cl, PhS (17). 4 TPs detected in SW and WW samples by UHPLC-TOF-MS and UHPLC-QqQ- MS	[22]
Díaz (2013)	UHPLC	Pharmaceuticals, illicit drugs and TPs (150 organic pollutants)	T	SW, GW, WW	Qualitative validation	[121]
Masiá (2013)	UHPLC	Pharmaceuticals, illicit drugs and other emerging	T, S	WW, SW	Quantification (42 pesticides by QqQ), confirmation (QqQ and QTOF) and	[75]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
		contaminants (250 T + 1100 S)			identification (QTOF)	
Vergeynst (2013)	UHPLC	Pharmaceuticals (69)	T	SW	Screening, quantification and confirmation	[70]
Boix (2014a)	UHPLC	Omeprazole metabolites (24)	T	WW, SW	Retrospective screening of metabolites previously identified in metabolism study. 9 metabolites detected in samples	[23]
Boix (2014b)	UHPLC	THC-COOH (Cannabis metabolite)	D	SW	Elucidation of Hy, Cl and PhS TPs (19). 8 TPs detected in SW and WW samples by UHPLC-QqQ-MS	[29]
Rodriguez- Alvarez (2014)	HPLC	Ethyl sulfate (biomarker for ethanol tracing)	T	WW	Quantification and confirmation (QqQ <i>versus</i> QToF)	[68]

^a T, Target; S, Suspect; NT, Non-Target; D, Degradation study.

^b HPLC, High-performance liquid chromatography, distilled, Ultrapure or demineralized water; SW, Surface water; WW, Wastewater; GW, Groundwater; DW, Drinking water; SS, Sewage sludge; A, Any artificial water matrix; MW, Marine environment; Sed, Sediment; n.d., Not clearly defined.

^c PhN, Photolysis with natural sunlight; PhS, Photolysis with simulated sunlight; Bd, Biodegradation; Hy, Hydrolysis; Pc, Photocatalysis; Cl, Chlorination; Oz, Ozonation.

^d Only TOF instrument used.

Table 2: Literature overview on LC-LTQ Orbitrap-MS applications

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
Hogenboom (2009)	HPLC	Pharmaceuticals, illicit drugs and other emerging contaminants (3000 T)	T, NT	SW, WW, GW, DW	Screening, quantification and identification	[118]
Hollender (2009)	HPLC	Emerging contaminants (220 T, including 77 pharmaceuticals); known oxidation TPs of 5 pharmaceuticals (S)	T,S	WW	Screening and removal efficiency for 220 pollutants after Oz degradation. Screening of known oxidation products of 5 pharmaceuticals	[85]
Kern (2009)	HPLC	Plausible TPs assembled using computer-aided prediction (UM-PPS) of pesticides (24) + biocides (7) + pharmaceuticals (21). TPs reported in literature of pharmaceuticals (21) and pesticides (31). Total: 1794 proposed TPs (890 from pharmaceuticals)	S	SW	19 TPs identified in SW (7 from pharmaceuticals) → 12 confirmed with reference standards	[28]
Helbling (2010)	HPLC	Plausible TPs assembled using computer-aided prediction (UM-PPS) of pharmaceuticals (6) + pesticides (6). Degradation study	S, D	WW	26 TPs identified after Bd (13 from pharmaceuticals). UM-PPS predicted the structures of 21 TPs. Elucidation resulted in 26 TPs	[84]
Kern (2010)	HPLC	Plausible TPs assembled using computer-aided prediction (UM-	S	WW	12 TPs identified after Bd. Application to real samples	[27]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
		PPS) + TPs reported in literature (7)			(quantification by QqQ)	
Bagnati (2011)	HPLC	Cocaine and benzoylecgonine	T	WW, SW	Evaluation of HRMS capabilities: screening, quantification and confirmation	[122]
Pinhancos (2011)	UHPLC	Pharmaceuticals (8) + metabolite of caffeine	T	DW	Screening, quantification and confirmation	[48]
Wille (2011)	UHPLC ^d	Pharmaceuticals (16) and pesticides (13)	T	MW	Screening, quantification and confirmation	[49]
Bijlsma (2012)	HPLC	Illicit drugs and metabolites (24)	T	WW	Quantification, confirmation and removal efficiency	[77]
Cahill (2012)	HPLC	Pharmaceuticals (5) and pesticides (4)	T	WW	Screening, quantification and confirmation	[78]
Calza (2012a)	HPLC	Clarithromycin and carbamazepine	D	HPLC	Elucidation of Pc TPs (28 from carbamazepine and 29 for clarithromycin)	[31]
Calza (2012b)	HPLC	Lincomycin	D	HPLC	Elucidation of Pc TPs(21)	[32]
de Jongh (2012)	HPLC	Pharmaceuticals and TPs (26)	T	SW, DW, GW	Screening and quantification	[79]
ter Laak (2012)	HPLC	Pharmaceuticals, illicit drugs, personal care products and other organic pollutants (635)	T	GW	Screening and confirmation	[82]
Bijlsma (2013)	HPLC	Illicit drugs and metabolites (24 T + 2 S)	T, S	WW	Quantification and confirmation.	[20]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
Calza (2013)	HPLC	Lincomycin, clarithromycin and carbamazepine	D	SW	Elucidation of PhS TPs (19 from lincomycin, 21 from carbamazepine and 6 from clarithromycin). 13 TPs detected in samples	[99]
Chiaia-Hernandez (2013)	HPLC	Pharmaceuticals, personal-care products and other organic pollutants (>180 T + 80 S)	T, S	Sed	Screening and identification (APPI and ESI)	[81]
van der Aa (2013)	HPLC	Illicit drugs and metabolites (34)	T	DW, SW, WW	Screening and quantification (incl. QqQ, Orbitrap) from different research groups	[123]
Rodayan (2014)	UHPLC	Illicit drugs and TPs (7)	D	WW	Elucidation of Oz TPs (10). Removal efficiency	[50]
Emke (2014)	HPLC	Amphetamine and MDMA	T	WW	Enantiomeric profiling. Quantification and confirmation	[83]
Hug (2014)	HPLC	Novel micropollutants (98 T + 2160 S)	T, S, NT	WW	Screening and identification	[26]
Kosma (2014)	HPLC	Pharmaceuticals and personal-care products (18)	T, S	WW	Confirmation of positives detected by LC-MS. Removal efficiency by LC-MS. Identification of trimethoprim TPs (2).	[80]
Schymanski (2014)	HPLC ^c	Polar organic contaminants (364 T + ~180 S)	T, S, NT	WW	Screening and identification	[119]

Author (year)	LC	Compound (no.) ^a	T/S/NT/D ^a	Matrix ^b	Observations ^c	Ref.
Zonja (2014)	HPLC	Zanamivir	D	A, SW	Identification of TPs after PhN and PhS (4). 1 TP confirmed with reference standard	[51]

^a T, Target; S, Suspect; NT, Non-target; D, Degradation study.

^b HPLC, High-performance liquid chromatography; distilled, Ultrapure or demineralized water; SW, Surface water; WW, Wastewater; GW, Groundwater; DW, Drinking water; SS, Sewage sludge; A, any artificial water matrix; MW, marine environment; Sed, Sediment.

^c PhN, Photolysis with natural sunlight; PhS, Photolysis with simulated sunlight; Bd, Biodegradation; Hy, Hydrolysis; Pc, Photocatalysis; Cl, Chlorination; Oz, Ozonation.

^d Exactive instrument used.

^e Q Exactive instrument used.

4.1. Screening of pharmaceuticals and illicit drugs in waters

LC-HRMS has extraordinary potential for screening PIDs in waters. The most efficient, rapid way is to perform target screening of a large number of compounds on the basis of large (in-house) databases. In this target approach, theoretical exact masses of analytes are extracted from full-spectrum acquisition data, reconstructing accurate-mass chromatograms, where the presence of analytes in the samples can be depicted as a chromatographic peak. Mass accuracy is critical for identification purposes. Normally, mass errors below 5 ppm are observed in routine analysis with the new instruments commercially available. The accurate mass of the (de)protonated molecule (on occasions adducts), the information on characteristic fragment ion(s), the isotopic pattern, and retention time matching with reference standards enable unambiguous identification of PIDs in environmental samples.

4.1.1. (Q)TOF-MS applications

TOF and/or hybrid QTOF analyzers are most commonly used in this field. The first works dealing with the analysis of pharmaceuticals in the environment by LC-QTOF-MS were reported in 2003. Marchese et al. [52] compared the potential of QqQ and QTOF for the LC-MS determination of five pharmaceuticals in river-water samples. According to the authors, the selectivity was much better on QTOF than on QqQ because of the high resolving power of the TOF analyzer, permitting high-accuracy fragment-ion selection and minimizing interferences from environmental matrices. In the same year, Ferrer et al. showed the benefits of the enhanced resolving power of LC-QTOF-MS when investigating polar organic contaminants in complex environmental matrices, as isobaric interferences could be separated from the analyte signals [53] and [54].

After these pioneering works, the capability of (Q)TOF-MS for screening, quantification and/or confirmation of pharmaceuticals in water samples has been regularly compared to that of LC-MS/MS with QqQ [11], [55], [56] and [57], and that resulted in a general agreement on the strong points and the weak points of these techniques. The elevated mass resolution and selectivity of TOF instruments diminish the problem of isobaric interferences [55], making TOF-MS very useful for not only identification and confirmation purposes, but also studying metabolic routes and degradation pathways [57]. The main limitation of TOF instruments has traditionally been lower sensitivity than QqQ instruments, making QqQ more appropriate for analysis at low ng L^{-1} concentrations of analytes [11]. In terms of dynamic range and limits of detection (LODs), the quantitative performance of QqQ instruments still presents better features than TOF-MS [55].

Several reviews were published in the past decade dealing with the use of LC-HRMS for the determination of PIDs in environmental analysis. We recommend reading of these reviews by researchers interested in this field [2], [3], [10], [13], [18], [58], [59], [60], [61], [62] and [63].

QTOF-MS has been used for unequivocal confirmation of the identity of pharmaceuticals previously detected and quantified by LC-MS/MS in SRM mode using hybrid quadrupole linear ion trap (QTRAP) [64] or QqQ [65]. The possibility of performing retrospective analysis has allowed the revision of recorded chromatograms for new compounds, metabolites or TPs in the samples, increasing the scope of the method in monitoring programs [21] and [64]. LC-MS/MS and LC-(Q)TOF-MS can be seen as complementary techniques. On the one hand, LC-MS/MS is the first choice for quantification in pre-target analysis, due to its good sensitivity and precision. On the other hand, QTOF provides accurate-mass measurements, and is ideal for post-target screening and confirmation. The most suitable strategy seems to be automated screening and identification by LC-(Q)TOF, followed by quantification by LC-MS/MS [65].

The use of (Q)TOF instruments for quantification of target analytes in environmental samples has been rather limited; however, in the past few years, there was increased use of quantitative analysis, since the limitations of lower sensitivity and linear dynamic range were mostly solved. This improvement contributed to extending the use of LC-(Q)TOF-MS to quantitation of target pharmaceuticals [66] and [67] and drugs [35] or alcohol biomarkers [68] in water.

The qualitative field is undoubtedly where HRMS can take full advantage of its capabilities derived from sensitive accurate-mass, full-spectrum acquisition. These capabilities offer the possibility of investigating the presence of compounds once sample analysis has been performed and MS data acquired (i.e., without pre-selection of analytes), and are especially suitable for screening/identification of contaminants and confirmation of presumed positive samples reported by other techniques.

LC-(Q)TOF-MS has been applied successfully to the investigation of pharmaceuticals belonging to different therapeutic classes in surface waters [55], [69] and [70], seawater [67] and wastewaters [55], [66], [71] and [72],

and also illicit drugs and their metabolites in wastewaters [35] and [73]. As a consequence of the intrinsic characteristics of HRMS analyzers, their use enables screening for a large number of contaminants with high sensitivity within one run, with the obvious restrictions derived from the chromatographic and ionization processes in LC-MS and from sample pre-treatment. LC-(Q)TOF-MS instruments have been applied for comprehensive screening of many pollutants of different chemical families, including PIDs [36], [64], [73] and [74] in water samples. In some cases, LC-(Q)TOF-MS has also been used as a complementary tool for target LC-QqQ-MS-based methods allowing the detection of analytes different to those selected in previous QqQ methods [75].

The easy review step and the useful information provided by TOF-MS gives high confidence in the identification of the compounds detected, even without reference standards being available. In these cases, tentative identification may be possible based on the presence of the (de)protonated molecule (i.e., when a chromatographic peak is detected at its accurate mass). Subsequently, the collision-induced dissociation (CID) fragments/product ions (or characteristic isotopic ions) are then evaluated [16]. For this purpose, different possibilities are available, such as comparing experimental MS(/MS) spectra or the main fragment ions with those reported in the literature (massbank, or METLIN public library), or justifying the accurate-mass fragments taking into account the structure of the molecule. For the latter approach, the use of specialized software (e.g., MassFragment) can help.

As an example, Fig. 3 illustrates the detection and the tentative identification of veterinary anthelmintic drug levamisole, recently also recognized as an adulterant in cocaine, in an effluent wastewater by using UHPLC-QTOF-MS, operating in MS^E mode. The protonated molecule of levamisole was detected in the LE function (Fig. 3a, bottom), with 3.4-ppm mass error. As the reference standard was not available at the laboratory, the accurate mass of the fragment ions was justified using the MassFragment software (Waters) in order to advance towards a tentative identification. To minimize spectral interferences that would complicate the identification process, recognizing which ions are and which are not fragments became essential. To this end, UHPLC was valuable for choosing perfectly co-eluting ions (see chromatographic peak at 3.86 min, Fig. 3b). The elemental composition for the two fragments detected in the HE function (Fig. 3a, top) (m/z 178.0689 and 123.0267) was calculated, obtaining errors less than 1 ppm in relation to the theoretical exact masses predicted. In addition to all information available, tentative identification of levamisole was supported by the MS/MS product ions reported in the literature, where the two fragments (m/z 178 and m/z 123) observed had been previously reported using an LTQ-Orbitrap with a resolution of 7,500 [76]. The final acquisition of the reference standard allowed the ultimate confirmation of this compound in the wastewater sample.

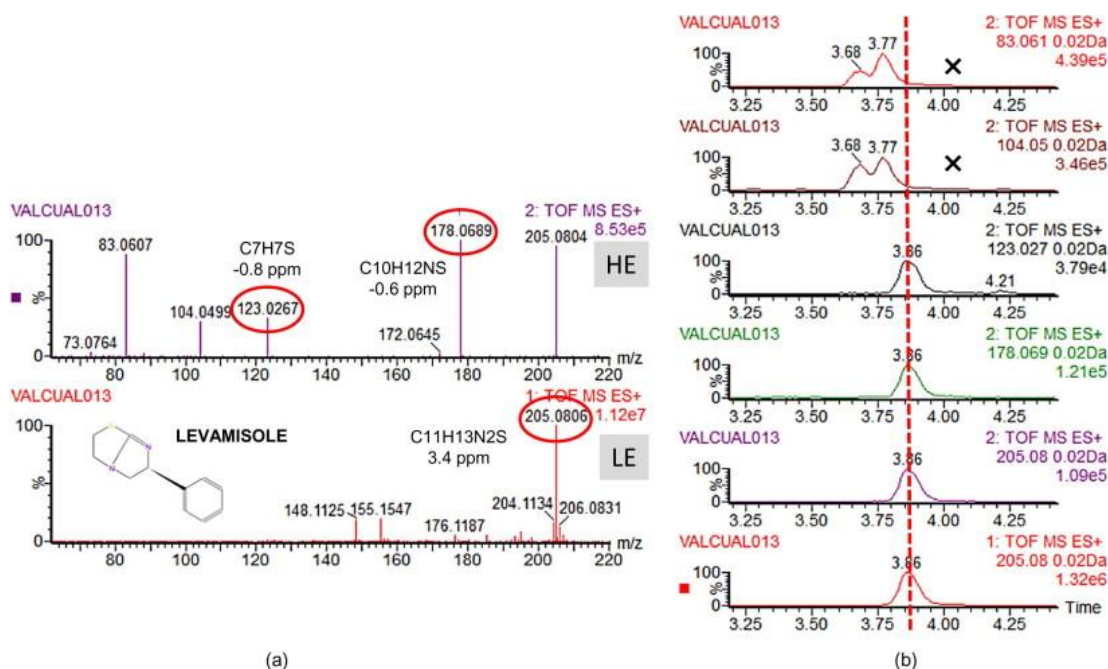


Fig.3: Detection and identification of levamisole by UHPLC-QTOF-MS in effluent wastewater (the reference standard was not available at the laboratory at the time of analysis): (a) Low-energy (LE) (bottom) and high-energy (HE) (top) spectra of the compound eluting at 3.86 min. (b) Extracted-ion chromatograms (0.02 Da mass width) for the protonated molecule in LE function and different fragment ions in HE function. (×) indicates that this ion is not related to levamisole

The interesting possibilities offered by hybrid QTOF-MS, including the option of working in MS^E mode, have led to a dramatic increase in the number of contaminants being included in the search, with more than 1000 compounds included in some cases [16] and [75] by applying suspect-screening approaches. This opens up a new scenario in screening strategies, favoring a wider, more realistic overview when investigating organic contaminants or residues in different applied fields.

4.1.2. (LTQ) Orbitrap-MS applications

The development of the Orbitrap analyzer in 2003 spurred researchers to explore its potential for accurate-mass screening, identification and quantification of PIDs in wastewaters [77] and [78], surface waters [79] and drinking waters [48] and [79] with good recoveries and RSD values. Similarly to QTOF, Orbitrap has also been used for the confirmation of PIDs previously found by other techniques [80] and for wide-scope target screening. A multi-residue method has been developed for target and suspect screening (i.e., compounds that did not form a part of the original target-screening list but were expected to be present in samples) of more than 180 organic contaminants, including pharmaceuticals, in lake sediments [81]. Orbitrap also enables retrospective analysis of the full-scan data, without the need for additional analysis [20], [48] and [49]. Assessing steps included retention-time prediction, isotope patterns, ionization efficiency and fragmentation pattern. Product ions of the MS/MS spectrum of a suspect compound were compared with the spectrum of a reference standard or with a predicted fragmentation pattern. This allowed the tentative identification of TPs of triclosan

and triclocarban. In another work, 42 groundwater samples were screened for 249 known chemicals and 386 unidentified chemicals (i.e., accurate masses and retention times) [82]. Nearly 400 chemicals were observed in the samples, of which 82 were known and more than 300 were of unknown identity.

Emke et al. [83] applied enantiomeric profiling in verifying sources of MDMA and amphetamine present in Dutch wastewater. The results from Orbitrap analysis showed that MDMA was usually present in wastewater due to its consumption as MDMA enriched with the R(-)-enantiomer. The high mass loads of racemic MDMA detected during a sampling campaign in The Netherlands therefore seemed to proceed from the direct disposal of unused MDMA, possibly as the result of a police raid at a nearby illegal production facility. HPLC separation was successfully performed using Chiral-CBH column, 100×2 mm, $5 \mu\text{m}$. Unfortunately, this approach is not yet feasible for UHPLC, due to the lack of chiral sub- $2\text{-}\mu\text{m}$ stationary phases.

4.2. Screening metabolites/transformation products in water samples

After human or animal consumption, PIDs or veterinary drugs may be excreted unchanged and/or as free or conjugated metabolites. Some of these compounds are not completely removed during wastewater treatment and may finally reach surface waters and even groundwaters. Consequently, several authors have included the main metabolites and TPs among the compounds monitored [21], but only a few works have focused on screening a notable number of TPs in aquatic compartments. Different approaches have been proposed in the literature for the investigation of metabolites and TPs, as shown below.

4.2.1. (Q)TOF-MS applications

Hernández et al. [21] studied the presence of 160 pharmaceutical metabolites in wastewater samples that had previously been analyzed for only parent compounds using LC-QTOF-MS (MS^E mode). The compounds investigated were selected after extensive search for TPs and metabolites reported in the literature and from lists of commercially available reference standards. As a first step, the presence of the metabolite ion (typically $[\text{M}+\text{H}]^+$ or $[\text{M}-\text{H}]^-$) at its accurate mass was evaluated. Different strategies were applied in the tentative identification:

- when the reference standard of the parent compound was available, and therefore accurate-mass fragment ions of the pharmaceutical were known, the accurate-mass fragments of the suspected metabolite were predicted, taking into account the structural differences between both molecules; and,
- when fragmentation information on the parent was not available, chemical structures of the accurate-mass fragments were proposed using specialized software (MassFragment).

In this case, information reported in the literature on product ions for the suspect metabolite or for the parent compound was essential (based on measurements of nominal mass or accurate mass). Following this strategy, several metabolites, such as clopidogrel carboxylic acid and N-desmethyl clarithromycin, were tentatively identified and subsequently confirmed after acquisition of the reference standards.

In a recent work, Boix et al. [23] investigated the presence of 24 omeprazole metabolites (identified in previous excretion tests performed with healthy volunteers) in surface water and effluent wastewater by UHPLC-QTOF-MS and UHPLC-MS/MS. Up to nine metabolites were detected, the most frequent being an omeprazole isomer, which presented the same exact mass (m/z 346.1225), and also shared a major common fragment at m/z 198.0589. By contrast, parent omeprazole was never detected in any of the water samples. The authors emphasized that monitoring the presence of omeprazole in the aquatic environment should focus on the main metabolites and some of the major TPs identified in laboratory experiments, instead of the parent compound.

4.2.2. (LTQ) Orbitrap-MS applications

Another strategy makes use of target lists of plausible TPs assembled using the University of Minnesota Pathway Prediction System (UM-PPS) for the computer-aided prediction of products of microbial metabolism and of metabolites and TPs reported in the literature. In this way, Kern et al. screened for potential biotransformation-product structures of pharmaceuticals in sludge-seeded batch reactors [27] or surface water [28] using HRMS (LTQ Orbitrap). These authors developed an identification procedure to examine a large number of proposed TPs efficiently for tentative identification without reference standards [28]. The procedure was based on a series of steps, with each step reducing the number of potential false-positive findings:

- (a) mass error of the measurements was <5 ppm for all compounds;
- (b) positive peak findings from the extracted chromatogram were discarded if the peak intensities in the extracted chromatogram were $<10^5$ (arbitrary units), or if a peak at similar retention and similar intensity was found in the blank sample;
- (c) experimental gradient retention times of the target TPs were tested against an upper plausible limit;
- (d) the isotope pattern of the HRMS spectra was evaluated to confirm a target molecular formula;
- (e) peaks with reasonable retention times and isotopic patterns were further checked for plausibility of a proposed TP in positive or negative ionization mode (e.g., compounds containing amino but no acidic groups were considered analyzable in positive-ionization mode only, whereas strong acids or sulfonates were expected to be detectable in negative mode); and,
- (f) to confirm further the presence of the target compounds in the samples, HR-MS/MS was used at higher energy collision-dissociation settings.

Altogether, 19 TPs were identified [28], including some rarely reported TPs, such as biotransformation products of pharmaceuticals venlafaxine and verapamil.

In another work, Helbling et al. [84] performed a preliminary identification of TPs formed within a biotransformation-test system using HRMS with two independent post-acquisition data-processing approaches:

- target screening at the exact masses of plausible TPs proposed by the UM-PPS for 12 selected parent compounds; and,
- screening for compound masses formed during biodegradation experiments based on background subtraction and exact mass filtering.

With this strategy, previously unreported microbial TPs were tentatively identified for several pharmaceuticals.

Hollender et al. [85] evaluated the removal efficiency of 220 micropollutants at a municipal WWTP upgraded with post-ozonation treatment followed by sand filtration. Afterwards, screening was conducted for known ozonation TPs of diclofenac, carbamazepine, and sulfamethoxazole, and possible oxidation products of benzotriazole and atenolol by LC-HRMS (Orbitrap).

4.3. Degradation/transformation under laboratory-controlled conditions

Once PIDs enter the WWTP, they can be completely mineralized, transformed into metabolites/TPs, adsorbed onto the solid phase (e.g., sewage sludge), or pass through the WWTP unaltered. Similarly, once released into the environment via the discharge of treated wastewater, PIDs are subjected to different processes, such as hydrolysis and photo-degradation by natural sunlight. In order to evaluate the fate of PIDs in the water cycle, the removal of the parent compounds and metabolites in the WWTP must be taken into account, but also the possible formation of TPs. The identification of TPs is complicated, but important, not only to provide a comprehensive risk assessment on drug residues in the environment (TPs can be equally or even more toxic and dangerous than the parent pollutants), but also to design improved treatment technologies for organic contaminants. From the different possibilities currently available, LC-HRMS has gained popularity and has become one of the techniques preferred for elucidating TPs in the environment. Several reviews have focused on the capabilities and the potential of MS techniques to determine pharmaceutical-degradation products [4], [5] and [86], although very little was included on the use of HRMS, surely due to the time-frame reviewed.

The process of identification and elucidation of unknown TPs is arduous, and commonly starts with laboratory experiments to study the formation of TPs under controlled conditions simulating those that occur in the environment. The most feasible strategy relies on MS measurement of their accurate molecular mass and subsequent determination of the empirical formula, considering low mass errors (e.g., <5 ppm). The maximum and minimum parameters considered in the elemental composition calculation may be restricted on the basis of the structure of the parent compound, and the number of halogen atoms selected according to the observed isotopic pattern. Further accurate MS/MS measurement and observation of characteristic fragmentation pattern of the precursor ion provide valuable information for identifying TPs. In some cases, HRMS alone is

insufficient to identify the exact position of transformation, to differentiate isomers, or to provide the precise structure of TPs. Thus, LC-HRMS can be combined with other techniques, such as nuclear magnetic resonance (NMR), or hydrogen/deuterium-exchange (H/D-exchange) for the final characterization of TP candidates of PIDs [87]. Although these techniques have little applicability in the environmental field, as they require high concentrations of analyte for measurement, they might be useful in laboratory experiments, where higher concentrations can be used.

Considerable time and effort are required to interrogate the intensive LC-HRMS data to identify unknowns. However, normally, the TPs are not complete unknowns as they hold a certain structural relationship with their parent compound. Some works only focus the research on the peaks visible in the total ion chromatogram (TIC), which implies that other invisible peaks, which may correspond to TPs, would be lost. To avoid this problem, detection can be improved by applying spectral and chromatographic search algorithms [e.g., MetaboLynx (Waters), Analyst/MetaboliteID (Applied Biosystems), Xcalibur/MetWorks (Thermo Fisher) or MassHunter (Agilent)]. Such algorithms search XICs for expected metabolites based on predicted or unpredicted molecular changes relative to the parent compound, and thus aid in the detection and identification of TPs, particularly those buried in spectral noise. The software compares mass spectral chromatograms of a control sample *versus* an analyte sample (i.e. metabolized, stressed or treated sample), and automates detection, identification and reporting of metabolites [88], [89] and [90].

PIDs can be degraded in natural environments or in WWTPs by both abiotic (advanced oxidation, photolysis and hydrolysis) or biotic (biodegradation) processes [87]. Different degradation/transformation experiments have been performed under laboratory-controlled conditions, trying to simulate the probable processes occurring in the aquatic environment and/or in WWTPs. Some representative examples are discussed below.

4.3.1. Hydrolysis experiments

Hydrolysis is a process occurring naturally in the environment. Pérez-Parada et al. [91] performed solely hydrolysis degradation experiments on amoxicillin in both alkaline and acidic aqueous media. Four compounds were identified as main TPs by LC-QTOF-MS. Data showed that although this antibiotic is not present as such in environmental samples, different TPs might occur. However, most papers reported on hydrolysis, studied this process together with other degradation experiments [22], [29] and [30].

4.3.2. Photodegradation experiments

Photodegradation by sunlight may constitute a relevant natural attenuation process for pollutant residues that have been discharged from WWTPs. It is one of the most important processes responsible for the degradation of contaminants in environmental systems [92], [93] and [94]. Understanding the degradation pathways is essential to predict the fate and the environmental impact of organic contaminants in waters. Photolytic

reactions lead to multiple reaction products that may be more toxic than the parent compound [95], retain the properties of the parent compound (e.g., antibiotic activity) [96], or lose antimicrobial activity and/or toxicity [95].

LC-TOF-MS has been used to elucidate the photo-TPs of several pharmaceuticals in water exposed to direct solar irradiation. Analytes studied included sulfa-drugs containing six-membered heterocyclic substituents (i.e., sulfamethazine, sulfamerazine, sulfadiazine, sulfachloropyridazine, and sulfadimethoxine) [88] and non-steroidal anti-inflammatory drug diclofenac [97]. For diclofenac, the authors combined LC-TOF-MS with GC-MS to detect and to identify a complete range of TPs, as both techniques provided complementary information. Similarly, the photochemical fate of prodrug enalapril and its active metabolite, enalaprilat [98], and sildenafil (Viagra) and its human metabolite, N-demethylsildenafil [90], have been investigated in aqueous media under the influence of a sunlight simulator. To postulate plausible chemical structures for several photo-TPs, accurate-mass measurements were combined with complementary data sets from different instruments:

- a QTRAP mass spectrometer exploiting its MS³ capabilities [98]; and,
- LC-APCI-QqQ MS and H/D-exchange experiments [90].

The aqueous environmental fate of pharmaceuticals has also been studied using LTQ Orbitrap MS. Calza et al. [31] and [32] investigated the degradation of antibiotics lincomycin and clarithromycin, and antiepileptic drug carbamazepine, whereas Zonja et al. [51] evaluated the phototransformation of antiviral zanamivir in surface waters using simulated and natural sunlight. Several species were formed under irradiation, and were characterized by evaluating MS and MSⁿ spectra. Fig. 4 shows an MS² spectrum of a dihydroxyl derivative of lincomycin. The presence of a product ion at m/z 407.1836, due to the loss of methanol from the N-hydroxymethyl moiety, combined with m/z 142.1219, indicates the existence of a hydroxyl group in the ortho position, and this allowed tentative structure elucidation. In the end, eight TPs from carbamazepine, three from clarithromycin, and two from lincomycin were detected in natural river water [99].

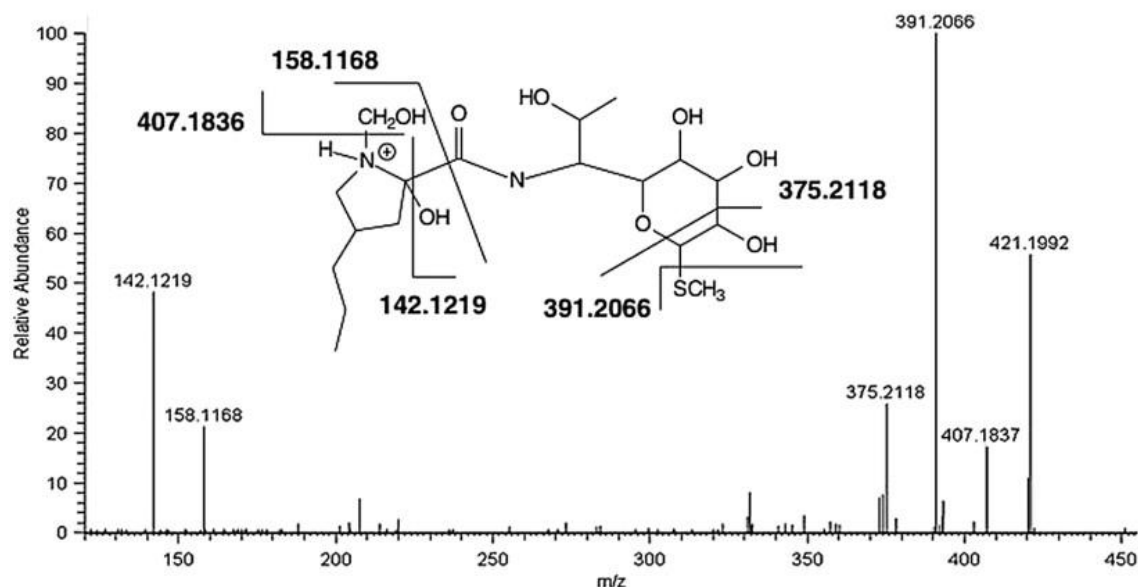


Fig. 4: MS2 spectrum of a dihydroxy derivative of lincomycin. {Reproduced from [32] with permission}.

Illicit drugs have also been subjected to degradation experiments, although they have been studied less than pharmaceuticals. Postigo et al. described the transformation of synthetic opioid methadone [100] and cocaine [101] in distilled water and simulated effluent wastewater after natural solar irradiation and two solar photocatalytic processes:

- heterogeneous photocatalysis with titanium dioxide; and,
- homogeneous photocatalysis by photo-Fenton.

Phototransformation intermediates generated during each treatment were investigated and characterized by UHPLC-QTOF-MS/MS. Identity confirmation was possible for some of them with the analysis of commercially available analytical standards.

The group of Hernández et al. performed laboratory-controlled degradation experiments for several PIDs in surface water. Hydrolysis, photo-degradation under ultraviolet (UV) irradiation and simulated sunlight, and chlorination experiments, were carried out. The TPs formed were investigated by LC-QTOF in MS^E mode. Studies were directed towards the degradation of cocaine and its main metabolite benzoylecgonine [30], THC-COOH (the main metabolite of cannabis and commonly selected as biomarker for the investigation of its consumption) [29] and omeprazole (one of the most consumed pharmaceuticals world-wide for treating gastric diseases) [22]. Six TPs from cocaine and 10 TPs from benzoylecgonine were tentatively identified. Regarding THC-COOH, one hydrolysis, eight chlorination, three UV photo-degradation and seven sunlight photo-degradation TPs were tentatively identified. In addition, 17 TPs were identified for omeprazole. In a subsequent step, samples of influent and effluent sewage water, and surface water, were retrospectively screened using UHPLC-QTOF-MS and UHPLC-QqQ MS for the TPs previously identified in the laboratory experiments. In

addition to some known compounds, several TPs that had not been reported in the literature were found in the samples, illustrating the usefulness of the degradation experiments performed.

As an illustrative example, Fig. 5 shows the detection and the identification of an omeprazole photo-TP (OTP 5) after 14 days of solar irradiation of spiked surface water. A chromatographic peak was observed at 5.68 min (Fig. 5a, bottom), while it was absent in the blank control sample (Fig. 5a, top). According to its accurate mass (m/z 330.1284, Fig. 5b, bottom), the elemental composition of the protonated molecule was assigned as $C_{17}H_{20}N_3O_2S^+$ (+2.4 ppm mass error), which would imply the loss of one oxygen atom from the omeprazole molecule. The fragment ions at m/z 297.1497 ($C_{17}H_{19}N_3O_2^+$, 6.7 ppm mass error) and m/z 282.1240 ($C_{16}H_{16}N_3O_2^+$, -1.1 ppm) (Fig. 5b, top) were assigned to thiol [\bullet SH] and subsequent methyl [\bullet CH $_3$] radical losses, respectively, and were related to the presence of a sulfide group [23]. Thus, an oxygen loss from the sulfur atom in the original omeprazole structure was suggested for this TP. This compound, identified in laboratory experiments, was later found in the water samples analyzed.

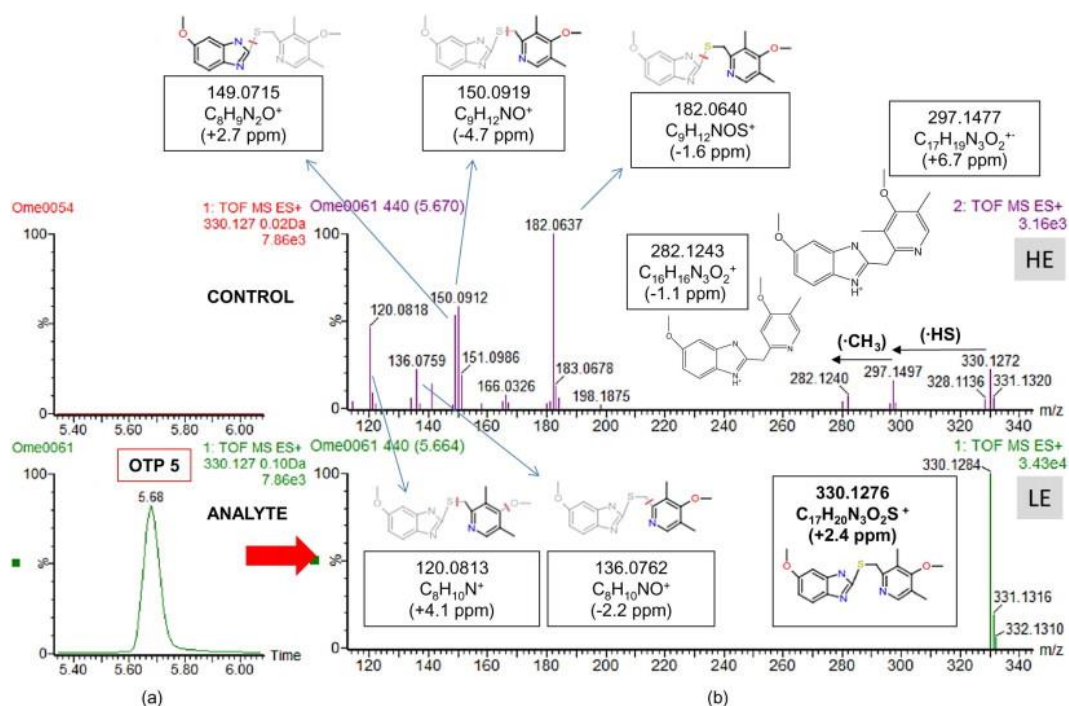


Fig. 5: Detection and identification of a TP of omeprazole (OTP 5) by LC-QTOF-MS (MS^E) resulting from photodegradation. (a) XIC at m/z 330.1276 (0.02-Da mass-window width) for analyte-sample (bottom) and control-sample (top), after 14 days of solar irradiation. (b) Low-energy (LE) (bottom) and high-energy (HE) (top) spectra and justification of fragment ions observed.

4.3.3. Advanced oxidation processes (AOP) experiments

Great efforts were made in recent years to develop additional (or alternative) processes for wastewater treatment. An interesting review on removal of pharmaceuticals from water was recently published, and is recommended for additional information on this issue [102]. Although ozonation leads to the elimination of

many organic compounds in aqueous solution, this is not necessarily accompanied by total mineralization [103]. In most cases, degradation by-products generated in the process persist after the parent compounds have been totally eliminated. Recently, it was reported that ozonation may release oxidation intermediates with enhanced toxicity for aquatic life [104] and [105]. This hypothesis highlights the need to characterize reaction mixtures in order to identify persistent and possibly toxic compounds.

Gómez-Ramos et al. [105] identified six TPs using LC-QTOF-MS after ozonation of antibiotic sulfamethoxazole under different operational conditions. In a recent study [50], several illicit drugs were subjected to ozonation to estimate their removal as a function of ozone dose and to identify the significant oxidation TPs. Once the potential TPs were identified by LC-LTQ-Orbitrap, their structures were elucidated by carrying out HRMSⁿ experiments (up to MS⁴).

4.3.4. Biodegradation experiments

The use of membrane bioreactor (MBR) technology for wastewater treatment presents advantages over the conventional activated sludge (CAS) process, such as enhanced biological performance, complete retention of solids, and smaller footprint [106].

To date, a number of abiotic PID degradation products have been identified, but published studies dealing with the structural elucidation of these degradates resulting from microbial transformation are still scarce. UHPLC-QTOF-MS has been used for screening and structural elucidation of biodegradation products, previously generated in small-scale laboratory batch reactors, of β -blocker atenolol and hypoglycemic agent glibenclamide [107], antimicrobial trimethoprim [89] and analgesic diclofenac [87]. In some cases, QTOF-MS has been combined with multiple-stage fragmentation studies and H/D-exchange experiments using IT-MS for final elucidation [89] or with HPLC-QqLIT-MS for confirmation of bio-TPs in wastewater samples [107]. For diclofenac, pre-processing based on isotopic cluster analysis was performed, since the parent compound contained two chlorine atoms in its chemical structure. In this way, the number of resulting peaks could be reduced to only those attributed to chlorinated diclofenac bio-TPs [87].

Helbling et al. [84] used HRMS Orbitrap data from a single bioreactor to identify compound masses formed during the biotransformation experiment. For each sample, a two-dimensional matrix of masses (m/z versus intensity) was found and their corresponding intensities were extracted. A script in Visual Basic was written to compare the mass-intensity matrices between $t = 0$ (sample withdrawn immediately after spiking with the compound of interest) and each $t > 0$ (samples withdrawn from the reactors at time points), and used a series of mass filters to reduce the number of extracted masses to a list of masses of candidate TPs. Candidate TP masses were further confirmed or rejected following manual inspection of MS spectra for the relative abundance of ¹³C and/or ³⁷Cl monoisotopic masses and/or adduct masses. Also, data-dependent MS/MS acquisitions were triggered when peaks were detected in full-scan mode at the exact masses of any parent

compound or candidate TP. Comparison of MS/MS fragments between each parent compound and TP further confirmed or rejected each candidate mass as a TP. Following this strategy, a dehydrogenation product of bezafibrate (not predicted by UM-PPS) was tentatively identified.

Degradation of sulfamethazine by white-rot fungus *Trametes versicolor* was also assessed. Four degradation intermediates produced by fungal cultures or purified laccase were identified using UHPLC–QTOF-MS [108].

4.4. Non-target analysis of water samples

4.4.1. Unbiased non-target analysis

The fact that HRMS provides full-spectrum acquisitions at accurate masses opens up the possibility of investigating non-target compounds in environmental samples. A genuine non-target analysis pursues the investigation (i.e., detection and identification or elucidation) of unknowns, which means that the analyst does not have any information on the compounds to be investigated (i.e., an unbiased analysis is pursued). The term “unknown” does not necessarily mean that the compound discovered in the analysis is new or not previously reported. It may be a very well-known compound that was not specifically searched for in analysis, so it was treated as an unknown [109].

It was early recognized that non-target analysis in environmental waters using LC-HRMS is a challenging task and successful identification of non-targets relies heavily on the (free) availability of chemical databases for finding candidates [25]. Two unidentified components in the pioneer work reported by Ibáñez et al. [25] were subsequently characterized thanks to the use of commercial chemical database SciFinder to find the candidate chemical structures [110].

A non-target analysis would require a visual inspection of the TIC to find potential components that might be present in the sample. Following this idea, Terzic et al. [111] developed a comprehensive analytical procedure based on UHPLC–QTOF-MS for the identification of non-target polar contaminants in aquatic sediments from a small water course greatly influenced by wastewater discharges from the pharmaceutical industry. The chromatograms, recorded in TIC mode, were systematically examined by manually generating mass spectra of each individual chromatographic peak. Only major peaks (>10% of the full scale intensity) were subjected to the identification process as non-target analytes. The first identification step was calculation of the possible elemental composition of the (de)protonated molecules from the mass spectra recorded, followed by a database search [using Chemspider, Merck Index and European chemical Substances Information System (ESIS), and an in-house database] for possible candidate compounds. For some contaminants, for which a reference standard was available, the information on accurate mass and information on retention time were sufficient for a reliable identification. For unknown compounds, the ultimate assignment of their identity was achieved by

MS/MS experiments, followed by comparing the observed fragmentation patterns with the main product ions expected for the selected candidates. Different pharmaceuticals were successfully identified and confirmed with reference compounds. However, eight prominent peaks remained unidentified.

One of the limitations of non-target analysis is that compounds of low abundance may not be apparent by visual inspection or, to the contrary, intense peaks may not necessarily be associated with a single component or “relevant” organic contaminants. Powerful software with chromatographic peak-deconvolution capabilities is required to detect multiple components and to produce pure spectra for each individual component. The non-target analysis workflow commonly starts with automated peak detection by exact-mass filtering from the chromatographic run, followed by assignment of an elemental formula to the exact mass of interest, and a subsequent search of plausible structures in available chemical databases for the elemental formula determined [15]. This is a laborious, difficult, time-consuming task when the unknown compounds are present at trace levels in complex-matrix samples.

4.4.2. Biased non-target analysis

An intermediate situation between target and non-target analysis is what some authors call a semi-non-target approach, based, for example, on the use of a chlorine mass-filter technique in accurate-mass and high-resolution systems [112]. Following this strategy, which only applies to the investigation of chlorine-containing compounds, antidepressant lamotrigine and its major metabolite (2-N-glucuronide) were identified in environmental water samples [112].

In non-target analysis, the use of large mass spectra libraries can notably facilitate the identification of unknowns. The main difficulty is the lack of LC-MS standardized libraries, as a consequence of the differences in the ionization efficiency between the existing interfaces together with variability in the results depending on the mobile-phase composition or the cone voltage applied. Under these circumstances, it is more convenient to build home-made libraries to simplify searching. Obviously, the higher the number of compounds included in the library, the wider the possibilities to detect as many contaminants as possible in the samples. With the software currently available, it is possible to build both empirical and/or theoretical libraries [16], [113] and [114]. Home-made empirical libraries offer the possibility to include fragmentation and retention-time information, which greatly aid the identification process. However, commonly, not many compounds are included, due to the need to inject reference standards. Instead, theoretical mass spectra libraries, based on a database with solely the molecular formulae, can easily be built including a large number of compounds. Díaz et al. [16] and [113] combined UHPLC-QTOF-MS (MS^E) with specialized software (ChromaLynx) for non-target screening of environmental water and wastewater. After chromatographic peak deconvolution, the software discriminated ions coming from organic compounds present in the sample from background ions. In a second step, a library search was performed to match the experimental deconvoluted mass spectra with the existing entries in the available libraries (theoretical and empirical). Finally, the formula from the library hit was

submitted to an elemental composition calculator and the most intense ions were either confirmed or rejected based on accurate-mass criteria and isotopic pattern.

In a similar way, Gómez-Ramos et al. [115] proposed a systematic strategy to simplify the identification of organic-contaminant TPs, based on characteristic fragmentation undergone by the parent compound during MS/MS fragmentation, and on the relationship with the transformations experienced by these chemicals in the environment or during water-treatment processes. A database containing accurate-mass information of 147 compounds (including pharmaceuticals, drugs and some relevant metabolites) and their main fragments generated by MS/MS was created using LC-QTOF-MS/MS. This database was applied to the tentative identification of TPs and related unexpected compounds in eight wastewater effluent samples. The approach comprised three stages:

- (a) automatic screening, consisting of extraction of compounds using a “molecular feature extraction” algorithm and database search;
- (b) identification of possible TPs (those reported compounds that yielded a good match in accurate mass but presented different retention times were initially considered as potential TPs or fragments of potential TPs); and,
- (c) confirmation by MS/MS analysis of the TPs tentatively identified in the previous step.

Eight degradation products, from pharmaceuticals acetaminophen, amoxicillin, carbamazepine, erythromycin and azithromycin, were tentatively identified. Three of them were confirmed by analysis of the corresponding analytical standards.

4.4.3. The “All-in-One” approach

In recent years, there has been a move towards the “All-in-One” approach, which involves the combination of suspect and non-target screening with (quantitative) target analysis. Thus, LC-QTOF-MS has been proposed for simultaneous quantitative screening of target analytes and qualitative analysis of non-target compounds [116]. For 10 target compounds, the use of specialized software allowed mass spectral data to be processed using LC retention time, accurate mass (error) and spectral purity score. Satisfactory results were obtained in terms of sensitivity and linearity. A second identification approach was presented based on library searching for compounds not included in the analytical method as target, but which were present in a commercial accurate-mass MS/MS library (including ~1200 pharmaceuticals and personal-care products). The spectra acquired were automatically searched by spectral comparison using the MS/MS library to assist in compound identification. Following this, ketorolac, trazodone, fluconazole, metformin and venlafaxine could be identified in river-water samples. Furthermore, other compounds not included in the library were also identified by screening the peaks of highest intensity in the samples and by analysis of the full-scan TOF-MS, isotope pattern and MS/MS spectra. This was the case for histaminergic loratadine.

Nurmi et al. [117] evaluated the performance of UHPLC-TOF-MS in the identification of emerging contaminants in spiked wastewater, using target and non-target analysis. The method was satisfactory for target analysis with information about retention times. For those compounds lacking retention time (suspect screening), a four-stage process for identification was developed: the number of candidate compounds was reduced by using the accurate mass of selected compounds in two steps (30 mDa *nw*-XIC (stage 1) and ± 5 mDa mass error (stage 2)), structure–property relationships [i.e., retention-time prediction (stage 3)] and isotope patterns of the analytes (stage 4). Non-target analysis was tested by applying a theoretical mass spectra library for a wastewater sample spiked with six pharmaceuticals. The results showed a large number of false identifications. In addition, manual processing of the data was considered laborious and ineffective. Finally, target analysis was applied to a real wastewater sample. The analysis revealed the presence of six compounds that were afterwards confirmed with reference standards. Three psycholeptics (nordiazepam, oxazepam and temazepam) could be tentatively identified.

Hogenboom et al. applied similar approaches but using an LTQ-Orbitrap mass spectrometer [118]. Full-scan accurate-mass measurements were compared with theoretical exact masses of known environmental microcontaminants and/or with their self-created accurate-mass MS and MSⁿ database, containing ~3000 water pollutants. The database included theoretical mass, retention time, retention time relative to two internal standards and elemental composition of product ions. For accurate masses not found in the mass database, a (possible) elemental composition was proposed, and searched in databases, such as Chemfinder or Chemspider, to find out whether the unknown compound was ever patented, studied or commercialized. MSⁿ measurements were performed to obtain information on the fragment ions generated in the LIT (nominal mass of product ions) within the same analysis. The structures found in the libraries were evaluated based on the fragmentation patterns observed in the product-ion spectra acquired simultaneously. Confirmation of identity was done by comparing retention time and fragmentation pattern to those of a reference standard. If no information on the reference standard was available, the theoretical log K_{ow} was calculated based on the proposed structure and compared to the retention time of the compound detected.

Very recently, Schymanski et al. [119] developed and applied a three-fold approach:

- (semi)-quantitative target analysis for 364 target compounds;
- suspect screening, demonstrated for sulfur-containing surfactants [evidence used to support (or reject) the tentative suspect identification included expected retention time behavior and interpretation of the MS/MS spectra]; and,
- non-target screening, to perform a comprehensive characterization of polar compounds in wastewater effluents.

A non-target mass list for each sample was compiled with the *enviMass* software. Isotope and adduct grouping of these non-target masses was subsequently performed using the “non-target” R package. Finally, non-target

identification was performed on selected masses from the top 30 most intense peaks. The program system MOLGEN-MS/MS (MOLEcular structure GENeration) was used to calculate molecular formulas from the exact mass and isotope patterns from the MS and MS/MS fragmentation information, if available. MetFusion was used to perform parallel searches of compound databases and spectral libraries and to perform *in silico* fragmentation of the candidate structures. The number of references per compound, retrieved from ChempSpider, was also used to rank candidates.

Systematic strategies with automated approaches are required to filter the suspect compounds to be searched and to select “relevant” peaks on which the identification efforts should focus. Hug et al. [26] established a screening procedure based on LC-HRMS to detect site-specific, suspected and formerly unknown contaminants in a WWTP effluent. First, the effluent was screened for the 98 target compounds using known retention time, exact mass and fragment ions. Some 15 target compounds were detected and quantified, including seven pharmaceuticals. Second, accurate-mass ion chromatograms and peak lists were generated from full-scan data using MZmine. For suspect screening, an initial list of 2160 site-specific and reported water contaminants was reduced to those amenable to LC-HRMS. Thus, only suspects containing at least one atom of nitrogen, phosphorus, oxygen, sulfur or any metal(loid) were considered ionizable by ESI. After searching MZmine peak lists for the exact masses of suspects (only those with intensity higher than 10^5 a.u.), presumably false positive detections were stepwise excluded by retention-time prediction, evaluation of isotope patterns, ionization behavior, and HRMS/MS spectra. Finally, in non-target analysis, masses for identification were selected using the R package “non-target” based on distinctive isotope patterns (containing ^{37}Cl , ^{81}Br , ^{34}S or ^{15}N isotope peaks) and intensity ($>10^6$ a.u.). For the remaining masses, all possible formulae were calculated and checked using the Seven Golden Rules to reduce the number to one or two molecular formulae, which were searched in the ChempSpider database, generating candidate lists for each formula. Only candidates with >10 data sources were considered of commercial importance and necessitated an automated HRMS/MS evaluation using MetFrag software. To confirm tentative identifications of compounds further, deuterium-exchange experiments were also conducted. Six suspected and five non-target chemicals were identified, of which two had not been previously reported as environmental pollutants. However, another five non-target compounds could not be confirmed by the reference standard of the most likely candidate structure.

5. Conclusions and future trends

HRMS provides high-resolution, accurate-mass and high full-scan sensitivity and selectivity, making it very attractive for both target and non-target screening. Other advantages are the possibility of performing retrospective data examination and tentative identification of compounds when reference standards are unavailable. HRMS is also very powerful in the identification of degradation products or TPs in laboratory experiments and the elucidation of unknown compounds.

Recent developments in HRMS have allowed improvement of analytical methodologies applied in environmental MS. The search for the “All-in-One” method and instrument will continue in the coming years, as combining all desired features in just one method and instrument is an exciting issue: qualitative and quantitative analysis, with possibilities for structural elucidation of unknowns. Some recent configurations appearing in the market in the past few years (e.g., Triple TOF from AB Sciex, Xevo G2 QTOF from Waters, and Q Exactive from Thermo) are trying to meet this purpose: high sensitivity and selectivity, accurate mass, sufficient fragmentation to provide information on ion fragments, and satisfactory linear dynamic range. In the near future, rapid growth in HRMS applications will surely occur, in not only environmental research but also other fields, such as food safety, toxicology and doping-control analysis.

UHPLC has become a very popular technique, replacing HPLC separations in the most recently developed methods. It has been successfully coupled to MS(/MS) analyzers working at nominal mass and to HR-TOF-MS analyzers, with evident advantages deriving from its superior performance: short chromatographic runs, better sensitivity and more efficient separations. However, its coupling to Orbitrap is still problematic due to the low scan rates attainable when selecting high resolving power.

The expected developments of HRMS will surely relate to improvement in scan speed for Orbitrap, which will allow coupling to UHPLC and certainly increase the number of applications, and in mass resolving power for TOF. Advances in accurate-mass full-acquisition data processing using more powerful, user-friendly software are also likely. New hybrid analyzers incorporating ion-mobility spectrometry (IMS), before or between mass analyzers, are appearing in the market, adding an orthogonal separation mechanism to classical m/z ratio. IMS will probably improve identification of co-eluting isomers or confirmation of suspect compounds in complex samples.

Acknowledgements

This work was developed with the financial support provided by the *Plan Nacional de I+D+I, Ministerio de Economía y Competitividad* (Project Ref. CTQ2012-36189). The authors also acknowledge financial support from *Generalitat Valenciana* (Group of Excellence Prometeo 2009/054; Collaborative Research on Environment and Food Safety ISIC/2012/016).

Richard Bade acknowledges the European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement No. [Marie Curie-FP7-PEOPLE Grant #317205] for his Early Stage Researcher (ESR) contract.

References

- [1] S.D. Richardson Environmental mass spectrometry: emerging contaminants and current issues *Anal. Chem.*, 84 (2012), pp. 747–778
- [2] P. Vázquez-Roig, C. Blasco, Y. Picó Advances in the analysis of legal and illegal drugs in the aquatic environment *Trends Anal. Chem.*, 50 (2013), pp. 65–77
- [3] B. Petrie, E.J. McAdam, M.D. Scrimshaw, J.N. Lester, E. Cartmell Fate of drugs during wastewater treatment *Trends Anal. Chem.*, 49 (2013), pp. 145–159
- [4] T. Kosjek, E. Heath, M. Petrović, D. Barceló Mass spectrometry for identifying pharmaceutical biotransformation products in the environment *Trends Anal. Chem.*, 26 (2007), pp. 1076–1085
- [5] M. Petrovic, D. Barceló LC-MS for identifying photodegradation products of pharmaceuticals in the environment *Trends Anal. Chem.*, 26 (2007), pp. 486–493
- [6] A. Agüera, M.J. Martínez Bueno, A.R. Fernández-Alba New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters *Environ. Sci. Pollut. Res. Int.*, 20 (2013), pp. 3496–3515
- [7] S. Castiglioni, K.V. Thomas, B. Kasprzyk-Hordern, L. Vandam, P. Griffiths Testing wastewater to detect illicit drugs: state of the art, potential and research needs *Sci. Total Environ.*, 487 (2014), pp. 613–620
- [8] K.V. Thomas, L. Bijlsma, S. Castiglioni, A. Covaci, E. Emke, R. Grabic, et al. Comparing illicit drug use in 19 European cities through sewage analysis *Sci. Total Environ.*, 432 (2012), pp. 432–439
- [9] S. Castiglioni, L. Bijlsma, A. Covaci, E. Emke, F. Hernández, M. Reid, et al. Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers *Environ. Sci. Technol.*, 47 (2013), pp. 1452–1460
- [10] M. Petrovic, M. Farré, M. Lopez de Alda, S. Perez, C. Postigo, M. Köck, et al. Recent trends in the liquid chromatography–mass spectrometry analysis of organic contaminants in environmental samples *J. Chromatogr. A* (2010), pp. 4004–4017
- [11] O.J. Pozo, C. Guerrero, J.V. Sancho, M. Ibáñez, E. Pitarch, E. Hogendoorn, et al. Efficient approach for the reliable quantification and confirmation of antibiotics in water using on-line solid-phase extraction liquid chromatography/tandem mass spectrometry *J. Chromatogr. A*, 1103 (2006), pp. 83–93

- [12] D.R. Baker, B. Kasprzyk-Hordern Multi-residue determination of the sorption of illicit drugs and pharmaceuticals to wastewater suspended particulate matter using pressurised liquid extraction, solid phase extraction and liquid chromatography coupled with tandem mass spectrometry *J. Chromatogr. A*, 1218 (2011), pp. 8036–8059
- [13] M. Farré, M. Holčápek, P.J. Schoenmakers, L. Kantiani, M. Petrovic, S. Pérez, et al. Achievements and future trends in the analysis of emerging organic contaminants in environmental samples by mass spectrometry and bioanalytical techniques *J. Chromatogr. A*, 1259 (2012), pp. 86–99
- [14] E. Gracia-Lor, J.V. Sancho, F. Hernández Multi-class determination of around 50 pharmaceuticals, including 26 antibiotics, in environmental and wastewater samples by ultra-high performance liquid chromatography-tandem mass spectrometry *J. Chromatogr. A*, 1218 (2011), pp. 2264–2275
- [15] M. Krauss, H. Singer, J. Hollender LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns *Anal. Bioanal. Chem*, 397 (2010), pp. 943–951
- [16] R. Díaz, M. Ibáñez, J.V. Sancho, F. Hernández Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLS-QTOF-MS *Anal. Methods* (2012), pp. 196–209
- [17] R. Plumb, J. Castro-Pérez, J. Granger, I. Beattie, K. Joncour, A. Wright Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry *Rapid Commun. Mass Spectrom*, 18 (2004), pp. 2331–2337
- [18] F. Hernández, J.V. Sancho, M. Ibáñez, E. Abad, T. Portolés, L. Mattioli Current use of high-resolution mass spectrometry in the environmental sciences *Anal. Bioanal. Chem*, 403 (2012), pp. 1251–1264
- [19] A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer Comprehensive comparison of liquid chromatography selectivity as provided by two types of liquid chromatography detectors (high resolution mass spectrometry and tandem mass spectrometry): “where is the crossover point?” *Anal. Chim. Acta*, 673 (2010), pp. 60–72
- [20] L. Bijlsma, E. Emke, F. Hernández, P. de Voogt, F. Hernandez Performance of the linear ion trap Orbitrap mass analyzer for qualitative and quantitative analysis of drugs of abuse and relevant metabolites in sewage water *Anal. Chim. Acta*, 768 (2013), pp. 102–110
- [21] F. Hernández, M. Ibáñez, E. Gracia-Lor, J.V. Sancho Retrospective LC-QTOF-MS analysis searching for pharmaceutical metabolites in urban wastewater *J. Sep. Sci*, 34 (2011), pp. 3517–3526

- [22] C. Boix, M. Ibáñez, J.V. Sancho, W.M.A. Niessen, F. Hernández Investigating the presence of omeprazole in waters by liquid chromatography coupled to low and high resolution mass spectrometry: degradation experiments *J. Mass Spectrom*, 48 (2013), pp. 1091–1100
- [23] C. Boix, M. Ibáñez, T. Zamora, J.V. Sancho, W.M.A. Niessen, F. Hernández Identification of new omeprazole metabolites in wastewaters and surface waters *Sci. Total Environ*, 468–469 (2014), pp. 706–714
- [24] M. Ibáñez, J.V. Sancho, L. Bijlsma, A.L.N. van Nuijs, A. Covaci, F. Hernández Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances *Trends Anal. Chem*, 57 (2014), pp. 107–117
- [25] M. Ibáñez, J.V. Sancho, Ó.J. Pozo, W. Niessen, F. Hernández Use of quadrupole time-of-flight mass spectrometry in the elucidation of unknown compounds present in environmental water *Rapid Commun. Mass Spectrom*, 19 (2005), pp. 169–178
- [26] C. Hug, N. Ulrich, T. Schulze, W. Brack, M. Krauss Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening *Environ. Pollut*, 184C (2014), pp. 25–32
- [27] S. Kern, R. Baumgartner, D.E. Helbling, J. Hollender, H. Singer, M.J. Loos, et al. A tiered procedure for assessing the formation of biotransformation products of pharmaceuticals and biocides during activated sludge treatment *J. Environ. Monit*, 12 (2010), pp. 2100–2111
- [28] S. Kern, K. Fenner, H.P. Singer, R.P. Schwarzenbach, J. Hollender Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry *Environ. Sci. Technol*, 43 (2009), pp. 7039–7046
- [29] C. Boix, M. Ibáñez, L. Bijlsma, J.V. Sancho, F. Hernández Investigation of cannabis biomarkers and transformation products in waters by liquid chromatography coupled to time of flight and triple quadrupole mass spectrometry *Chemosphere*, 99 (2014), pp. 64–71
- [30] L. Bijlsma, C. Boix, W.M.A. Niessen, M. Ibáñez, J.V. Sancho, F. Hernández Investigation of degradation products of cocaine and benzoylecgonine in the aquatic environment *Sci. Total Environ*, 443 (2013), pp. 200–208
- [31] P. Calza, C. Medana, E. Padovano, V. Giancotti, C. Baiocchi Identification of the unknown transformation products derived from clarithromycin and carbamazepine using liquid chromatography/high-resolution mass spectrometry *Rapid Commun. Mass Spectrom*, 26 (2012), pp. 1687–1704

- [32] P. Calza, C. Medana, E. Padovano, F. Dal Bello, C. Baiocchi Identification of the unknown transformation products derived from lincomycin using LC-HRMS technique *J. Mass Spectrom*, 47 (2012), pp. 751–759
- [33] K.P. Bateman, J. Castro-perez, M. Wrona, J.P. Shockcor, K. Yu, R. Oballa, et al. MS E with mass defect filtering for in vitro and in vivo metabolite identification *Rapid Commun. Mass Spectrom*, 21 (2007), pp. 1485–1496
- [34] Y. Zhang, S.B. Ficarro, S. Li, J.A. Marto Optimized Orbitrap HCD for quantitative analysis of phosphopeptides *J. Am. Soc. Mass Spectrom*, 20 (2009), pp. 1425–1434
- [35] I. González-Mariño, J.B. Quintana, I. Rodriguez, M. González-Díez, R. Cela Screening and selective quantification of illicit drugs in wastewater by mixed-mode solid-phase extraction and quadrupole-time-of-flight liquid chromatography–mass spectrometry *Anal. Chem*, 84 (2012), pp. 1708–1717
- [36] I. Ferrer, E.M. Thurman Analysis of 100 pharmaceuticals and their degradates in water samples by liquid chromatography/quadrupole time-of-flight mass spectrometry *J. Chromatogr. A*, 1259 (2012), pp. 148–157
- [37] J. Sherma, D. John, F.H. Larkin UPLC: Ultra-performance liquid chromatography *J. AOAC Int*, 88 (2005), pp. 63A–67A
- [38] M. Mezcua, A. Agüera, J.L. Lliberia, M.A. Cortés, B. Bagó, A.R. Fernández-Alba Application of ultra performance liquid chromatography-tandem mass spectrometry to the analysis of priority pesticides in groundwater *J. Chromatogr. A*, 1109 (2006), pp. 222–227
- [39] C.C. Leandro, P. Hancock, R.J. Fussell, B.J. Keely Comparison of ultra-performance liquid chromatography and high-performance liquid chromatography for the determination of priority pesticides in baby foods by tandem quadrupole mass spectrometry *J. Chromatogr. A*, 1103 (2006), pp. 94–101
- [40] O.J. Pozo, M. Barreda, J.V. Sancho, F. Hernández, J. Ll Lliberia, M.A. Cortés, et al. Multiresidue pesticide analysis of fruits by ultra-performance liquid chromatography tandem mass spectrometry *Anal. Bioanal. Chem*, 389 (2007), pp. 1765–1771
- [41] B. Kasprzyk-Hordern, R.M. Dinsdale, A.J. Guwy Multi-residue method for the determination of basic/neutral pharmaceuticals and illicit drugs in surface water by solid-phase extraction and ultra performance liquid chromatography-positive electrospray ionisation tandem mass spectrometry *J. Chromatogr. A*, 1161 (2007), pp. 132–145

- [42] M. Gros, S. Rodríguez-Mozaz, D. Barceló Fast and comprehensive multi-residue analysis of a broad range of human and veterinary pharmaceuticals and some of their metabolites in surface and treated waters by ultra-high-performance liquid chromatography coupled to quadrupole-linear ion trap tandem J. Chromatogr. A, 1248 (2012), pp. 104–121
- [43] A.L.N. van Nuijs, S. Castiglioni, I. Tarcomnicu, C. Postigo, M. Lopez de Alda, H. Neels, et al. Illicit drug consumption estimations derived from wastewater analysis: a critical review Sci. Total Environ, 409 (2011), pp. 3564–3577
- [44] M. Hardman, A.A. Makarov Interfacing the orbitrap mass analyzer to an electrospray ion source Anal. Chem, 75 (2003), pp. 1699–1705
- [45] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R. Graham Cooks The Orbitrap: a new mass spectrometer J. Mass Spectrom, 40 (2005), pp. 430–443
- [46] M. Kellmann, H. Muenster, P. Zomer, H. Mol Full scan MS in comprehensive qualitative and quantitative residue analysis in food and feed matrices: how much resolving power is required? J. Am. Soc. Mass Spectrom, 20 (2009), pp. 1464–1476
- [47] A. Makarov, M. Scigelova Coupling liquid chromatography to Orbitrap mass spectrometry J. Chromatogr. A, 1217 (2010), pp. 3938–3945
- [48] R. Pinhancos, S. Maass, D.M. Ramanathan High-resolution mass spectrometry method for the detection, characterization and quantitation of pharmaceuticals in water J. Mass Spectrom, 46 (2011), pp. 1175–1181
- [49] K. Wille, M. Claessens, K. Rappé, E. Monteyne, C.R. Janssen, H.F. De Brabander, et al. Rapid quantification of pharmaceuticals and pesticides in passive samplers using ultra high performance liquid chromatography coupled to high resolution mass spectrometry J. Chromatogr. A, 1218 (2011), pp. 9162–9173
- [50] A. Rodayan, P.A. Segura, V. Yargeau Ozonation of wastewater: removal and transformation products of drugs of abuse Sci. Total Environ, 487 (2014), pp. 763–770
- [51] B. Zonja, C. Gonçalves, S. Pérez, A. Delgado, M. Petrovic, M.F. Alpendurada, et al. Evaluation of the phototransformation of the antiviral zanamivir in surface waters through identification of transformation products J. Hazard. Mater, 265 (2014), pp. 296–304

- [52] S. Marchese, A. Gentili, D. Perret, D. Ascenzo, F. Pastori Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of non-steroidal antiinflammatory drugs in surface water by liquid chromatography/tandem mass spectrometry *Rapid Commun. Mass Spectrom*, 17 (2003), pp. 879–886
- [53] I. Ferrer Liquid chromatography/time-of-flight/mass spectrometry (LC/TOF/MS) for the analysis of emerging contaminants *Trends Anal. Chem*, 22 (2003), pp. 750–756
- [54] I. Ferrer, E.M. Thurman Liquid chromatography/mass spectrometry, MS/MS and time of flight MS: analysis of emerging contaminants *ACS Symp. Ser*, 850 (2003), pp. 1–13
- [55] M. Farré, M. Gros, B. Hernández, M. Petrovic, P. Hancock, D. Barceló Analysis of biologically active compounds in water by ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry *Rapid Commun. Mass Spectrom*, 22 (2008), pp. 41–51
- [56] A.A.M. Stolker, W. Niesing, E.A. Hogendoorn, J.F.M. Versteegh, R. Fuchs, U.A.T. Brinkman Liquid chromatography with triple-quadrupole or quadrupole-time of flight mass spectrometry for screening and confirmation of residues of pharmaceuticals in water *Anal. Bioanal. Chem*, 378 (2004), pp. 955–963
- [57] A.A.M. Stolker, W. Niesing, R. Fuchs, R.J. Vreeken, W.M.A. Niessen, U.A.T. Brinkman Liquid chromatography with triple-quadrupole and quadrupole-time-of-flight mass spectrometry for the determination of micro-constituents – a comparison *Anal. Bioanal. Chem*, 378 (2004), pp. 1754–1761
- [58] M. Petrović, M.D. Hernando, M.S. Díaz-Cruz, D. Barceló Liquid chromatography–tandem mass spectrometry for the analysis of pharmaceutical residues in environmental samples: a review *J. Chromatogr. A*, 1067 (2005), pp. 1–14
- [59] M. Petrović, D. Barcelo Application of liquid chromatography / quadrupole time-of-flight mass spectrometry (LC-QqTOF-MS) *J. Mass Spectrom*, 41 (2006), pp. 1259–1267
- [60] D. Barceló, M. Petrovic Challenges and achievements of LC-MS in environmental analysis: 25 years on *Trends Anal. Chem*, 26 (2007), pp. 2–11
- [61] F. Hernández, J.V. Sancho, M. Ibáñez, C. Guerrero Antibiotic residue determination in environmental waters by LC-MS *Trends Anal. Chem*, 26 (2007), pp. 466–485
- [62] C.S. Wong, S.L. MacLeod JEM spotlight: recent advances in analysis of pharmaceuticals in the aquatic environment *J. Environ. Monit*, 11 (2009), pp. 923–936

- [63] I. Ferrer, M.E. Thurman Analysis of pharmaceuticals in drinking water, groundwater, surface water, and wastewater M. Petrovic, S. Perez, D. Barcelo (Eds.), *Comprehensive Analytical Chemistry*, vol. 62, Analysis, Removal, Effects and Risk of pharmaceuticals in the water cycle: occurrence and transformation in the environment (electronic book) (2013), pp. 91–128 Elsevier
- [64] M.J. Martínez Bueno, A. Agüera, M.J. Gómez, M.D. Hernando, J.F. García-Reyes, A.R. Fernandez-Alba Application of liquid chromatography/quadrupole-linear Ion trap mass spectrometry and time-of-flight mass spectrometry to the determination of pharmaceuticals and related contaminants in wastewater *Anal. Chem*, 79 (2007), pp. 9372–9384
- [65] R. López-Roldán, M.L. de Alda, M. Gros, M. Petrovic, J. Martín-Alonso, D. Barceló Advanced monitoring of pharmaceuticals and estrogens in the Llobregat River basin (Spain) by liquid chromatography–triple quadrupole-tandem mass spectrometry in combination with ultra performance liquid chromatography–time of flight-mass spectrometry *Chemosphere*, 80 (2010), pp. 1337–1344
- [66] H. Leknes, I.E. Sturtzel, C. Dye Environmental release of oseltamivir from a Norwegian sewage treatment plant during the 2009 influenza A (H1N1) pandemic *Sci. Total Environ*, 414 (2012), pp. 632–638
- [67] J. Magnér, M. Filipovic, T. Alsberg Application of a novel solid-phase-extraction sampler and ultra-performance liquid chromatography quadrupole-time-of-flight mass spectrometry for determination of pharmaceutical residues in surface sea water *Chemosphere*, 80 (2010), pp. 1255–1260
- [68] T. Rodríguez-Álvarez, R. Rodil, R. Celz, J.B. Quintana Ion-pair reversed-phase liquid chromatography-quadrupole-time-of-flight and triple-quadrupole-mass-spectrometry determination of ethyl sulfate in wastewater for alcohol consumption tracing *J. Chromatogr. A* (2014), pp. 35–42
- [69] M. Ibáñez, C. Guerrero, J.V. Sancho, F. Hernández Screening of antibiotics in surface and wastewater samples by ultra-high-pressure liquid chromatography coupled to hybrid quadrupole time-of-flight mass spectrometry *J. Chromatogr. A*, 1216 (2009), pp. 2529–2539
- [70] L. Vergeynst, H. Van Langenhove, P. Joos, K. Demeestre High-resolution time-of-flight mass spectrometry for suspect screening and target quantification of pharmaceuticals in river water *Commun. Agric. Appl. Biol. Sci*, 78 (2013), pp. 9–14
- [71] M. Petrovic, M. Gros, D. Barceló Multi-residue analysis of pharmaceuticals in wastewater by ultra-performance liquid chromatography-quadrupole-time-of-flight mass spectrometry *J. Chromatogr. A*, 1124 (2006), pp. 68–81

- [72] M. Lavén, T. Alsberg, Y. Yu, M. Adolfsson-Erici, H. Sun Serial mixed-mode cation- and anion-exchange solid-phase extraction for separation of basic, neutral and acidic pharmaceuticals in wastewater and analysis by high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry *J. Chromatogr. A*, 1216 (2009), pp. 49–62
- [73] F. Hernández, L. Bijlsma, J.V. Sancho, R. Díaz, M. Ibáñez Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography-quadrupole-time-of-flight-mass spectrometry *Anal. Chim. Acta*, 684 (2011), pp. 96–106
- [74] M.J. Gómez, M.M. Gómez-Ramos, O. Malato, M. Mezcua, A.R. Fernández-Alba Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography–quadrupole-time-of-flight mass spectrometry with an accurate-mass *J. Chromatogr. A*, 1217 (2010), pp. 7038–7054
- [75] A. Masiá, M. Ibáñez, C. Blasco, J.V. Sancho, Y. Picó, F. Hernández Combined use of liquid chromatography triple quadrupole mass spectrometry and liquid chromatography quadrupole time-of-flight mass spectrometry in systematic screening of pesticides and other contaminants in water samples *Anal. Chim. Acta*, 761 (2013), pp. 117–127
- [76] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, et al. MassBank: a public repository for sharing mass spectral data for life sciences *J. Mass Spectrom*, 45 (2010), pp. 703–714
- [77] L. Bijlsma, E. Emke, F. Hernández, P. de Voogt Investigation of drugs of abuse and relevant metabolites in Dutch sewage water by liquid chromatography coupled to high resolution mass spectrometry *Chemosphere*, 89 (2012), pp. 1399–1406
- [78] M.G. Cahill, B.A. Dineen, M.A. Stack, K.J. James A critical evaluation of liquid chromatography with hybrid linear ion trap-Orbitrap mass spectrometry for the determination of acidic contaminants in wastewater effluents *J. Chromatogr. A*, 1270 (2012), pp. 88–95
- [79] C.M. de Jongh, P.J.F. Kooij, P. de Voogt, T.L. ter Laak Screening and human health risk assessment of pharmaceuticals and their transformation products in Dutch surface waters and drinking water *Sci. Total Environ* (2012), pp. 70–77
- [80] C.I. Kosma, D.A. Lambropoulou, T.A. Albanis Investigation of PPCPs in wastewater treatment plants in Greece: occurrence, removal and environmental risk assessment *Sci. Total Environ*, 466–467 (2014), pp. 421–438

- [81] A.C. Chiaia-Hernandez, M. Krauss, J. Hollender Screening of lake sediments for emerging contaminants by liquid chromatography atmospheric pressure photoionization and electrospray ionization coupled to high resolution mass spectrometry *Environ. Sci. Technol*, 47 (2013), pp. 976–986
- [82] T.L. ter Laak, L.M. Puijker, J.A. van Leerdam, K.J. Raat, A. Kolkman, P. de Voogt, et al. Broad target chemical screening approach used as tool for rapid assessment of groundwater quality *Sci. Total Environ*, 427–428 (2012), pp. 308–313
- [83] E. Emke, S. Evans, B. Kasprzyk-Hordern, P. de Voogt Enantiomer profiling of high loads of amphetamine and MDMA in communal sewage: a Dutch perspective *Sci. Total Environ*, 487 (2014), pp. 666–672
- [84] D.E. Helbling, J. Hollender, H.-P.E. Kohler, H. Singer, K. Fenner High-throughput identification of microbial transformation products of organic micropollutants *Environ. Sci. Technol*, 44 (2010), pp. 6621–6627
- [85] J. Hollender, S.G. Zimmermann, S. Koepke, M. Krauss, C.S. McArdell, C. Ort, et al. Elimination of organic micropollutants in a municipal wastewater treatment plant upgraded with a full-scale post-ozonation followed by sand filtration *Environ. Sci. Technol*, 43 (2009), pp. 7862–7869
- [86] J. Radjenović, M. Petrović, D. Barceló Advanced mass spectrometric methods applied to the study of fate and removal of pharmaceuticals in wastewater treatment *Trends Anal. Chem*, 26 (2007), pp. 1132–1144
- [87] T. Kosjek, D. Žigon, B. Kralj, E. Heath The use of quadrupole-time-of-flight mass spectrometer for the elucidation of diclofenac biotransformation products in wastewater *J. Chromatogr. A*, 1215 (2008), pp. 57–63
- [88] A.L. Boreen, W.A. Arnold, K. McNeill Triplet-sensitized photodegradation of sulfa drugs containing six-membered heterocyclic groups: identification of an SO₂ extrusion photoproduct
Environ. Sci. Technol, 39 (2005), pp. 3630–3638
- [89] P. Eichhorn, P.L. Ferguson, S. Pérez, D.S. Aga Application of ion trap-MS with H/D exchange and QqTOF-MS in the identification of microbial degradates of trimethoprim in nitrifying activated sludge *Anal. Chem*, 77 (2005), pp. 4176–4184
- [90] P. Eichhorn, S. Pérez, J. Aceña, P. Gardinali, J.L. Abad, D. Barceló Identification of phototransformation products of sildenafil (Viagra) and its N-demethylated human metabolite under simulated sunlight *J. Mass Spectrom*, 47 (2012), pp. 701–711
- [91] A. Pérez-Parada, A. Agüera, M.D.M. Gómez-Ramos, J.F. García-Reyes, H. Heinzen, A.R. Fernández-Alba Behavior of amoxicillin in wastewater and river water: identification of its main transformation products by

liquid chromatography/electrospray quadrupole time-of-flight mass spectrometry *Rapid Commun. Mass Spectrom*, 25 (2011), pp. 731–742

[92] T.E. Doll, F.H. Frimmel Fate of pharmaceuticals – photodegradation by simulated solar UV-light *Chemosphere*, 52 (2003), pp. 1757–1769

[93] D.E. Latch, B.L. Stender, J.L. Packer, W.A. Arnold, K. McNeill Photochemical fate of pharmaceuticals in the environment: cimetidine and ranitidine *Environ. Sci. Technol*, 37 (2003), pp. 3342–3350

[94] A.L. Boreen, W.A. Arnold, K. McNeill Photodegradation of pharmaceuticals in the aquatic environment: a review *Aquat. Sci*, 65 (2003), pp. 320–341

[95] M. DellaGreca, A. Fiorentino, M.R. Iesce, M. Isidori, A. Nardelli, L. Previtiera, et al. Identification of phototransformation products of prednisone by sunlight: toxicity of the drug and its derivatives on aquatic organisms *Environ. Toxicol. Chem*, 22 (2003), pp. 534–539

[96] B. Halling-Sørensen, G. Sengeløv, J. Tjørnelund Toxicity of tetracyclines and tetracycline degradation products to environmentally relevant bacteria, including selected tetracycline-resistant bacteria *Arch. Environ. Contam. Toxicol*, 42 (2002), pp. 263–271

[97] A. Agüera, L.A. Pérez Estrada, I. Ferrer, E.M. Thurman, S. Malato, A.R. Fernández-Alba Application of time-of-flight mass spectrometry to the analysis of phototransformation products of diclofenac in water under natural sunlight *J. Mass Spectrom*, 40 (2005), pp. 908–915

[98] S. Pérez, P. Eichhorn, D. Barceló. Structural characterization of photodegradation products of enalapril and its metabolite enalaprilat obtained under simulated environmental conditions by hybrid quadrupole-linear ion trap-MS and quadrupole-time-of-flight-MS *Anal. Chem*, 79 (2007), pp. 8293–8300

[99] P. Calza, C. Medana, E. Padovano, V. Giancotti, C. Minero Fate of selected pharmaceuticals in river waters *Environ. Sci. Pollut. Res. Int*, 20 (2013), pp. 2262–2270

[100] C. Postigo, C. Sirtori, I. Oller, S. Malato, M.I. Maldonado, M. López de Alda, et al. Photolytic and photocatalytic transformation of methadone in aqueous solutions under solar irradiation: kinetics, characterization of major intermediate products and toxicity evaluation *Water Res*, 45 (2011), pp. 4815–4826

[101] C. Postigo, C. Sirtori, I. Oller, S. Malato, M.I. Maldonado, M. López de Alda, et al. Solar transformation and photocatalytic treatment of cocaine in water: kinetics, characterization of major intermediate products and toxicity evaluation *Appl. Catal. B Environ*, 104 (2011), pp. 37–48

- [102] J. Rivera-Utrilla, M. Sánchez-Polo, M.Á. Ferro-García, G. Prados-Joya, R. Ocampo-Pérez Pharmaceuticals as emerging contaminants and their removal from water. A review *Chemosphere*, 93 (2013), pp. 1268–1287
- [103] M. Klavarioti, D. Mantzavinos, D. Kassinos Removal of residual pharmaceuticals from aqueous systems by advanced oxidation processes *Environ. Int.*, 35 (2009), pp. 402–417
- [104] R.F. Dantas, M. Canterino, R. Marotta, C. Sans, S. Esplugas, R. Andreozzi Bezafibrate removal by means of ozonation: primary intermediates, kinetics, and toxicity assessment *Water Res.*, 41 (2007), pp. 2525–2532
- [105] M.D.M. Gómez-Ramos, M. Mezcuca, A. Agüera, A.R. Fernández-Alba, S. Gonzalo, A. Rodríguez, et al. Chemical and toxicological evolution of the antibiotic sulfamethoxazole under ozone treatment in water solution *J. Hazard. Mater.*, 192 (2011), pp. 18–25
- [106] J. Radjenović, M. Matošić, I. Mijatović, M. Petrović, D. Barceló Membrane Bioreactor (MBR) as an advanced wastewater treatment technology *Hdb. Env. Chem*, 5 (2008), pp. 37–101 in: D. Barceló, M. Petrovic (Editors), *Water Pollut.*, Springer Berlin Heidelberg, Berlin, Heidelberg
- [107] J. Radjenović, S. Pérez, M. Petrović, D. Barceló Identification and structural characterization of biodegradation products of atenolol and glibenclamide by liquid chromatography coupled to hybrid quadrupole time-of-flight and quadrupole ion trap mass spectrometry *J. Chromatogr. A*, 1210 (2008), pp. 142–153
- [108] M.J. García-Galán, C.E. Rodríguez-Rodríguez, T. Vicent, G. Caminal, M.S. Díaz-Cruz, D. Barceló Biodegradation of sulfamethazine by *Trametes versicolor*: removal from sewage sludge and identification of intermediate products by UPLC-QqTOF-MS *Sci. Total Environ.*, 409 (2011), pp. 5505–5512
- [109] F. Hernández, J.V. Sancho, M. Ibáñez, T. Portolés TOF and QTOF-MS for identifying unknown contaminants and degradation products in the environment R.A. Meyers (Ed.), *Encycl. Anal. Chem.*, John Wiley & Sons, Ltd., Chichester (2011)
- [110] A.H. Grange, M.C. Zumwalt, G.W. Sovocool Determination of ion and neutral loss compositions and deconvolution of product ion mass spectra using an orthogonal acceleration time-of-flight mass spectrometer and an ion correlation program *Rapid Commun. Mass Spectrom.*, 20 (2006), pp. 89–102
- [111] S. Terzic, M. Ahel Nontarget analysis of polar contaminants in freshwater sediments influenced by pharmaceutical industry using ultra-high-pressure liquid chromatography-quadrupole time-of-flight mass spectrometry *Environ. Pollut.*, 159 (2011), pp. 557–566

- [112] I. Ferrer, E.M. Thurman Identification of a new antidepressant and its glucuronide metabolite in water samples using liquid chromatography/quadrupole time-of-flight mass spectrometry *Anal. Chem*, 82 (2010), pp. 8161–8168
- [113] R. Díaz, M. Ibáñez, J.V. Sancho, F. Hernández Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/ hybrid quadrupole time-of-flight mass spectrometry *Rapid Commun. Mass Spectrom*, 25 (2011), pp. 355–369
- [114] M. Ibáñez, J.V. Sancho, D. McMillan, R. Rao, F. Hernández Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-flight mass spectrometry *Trends Anal. Chem*, 27 (2008), pp. 481–489
- [115] M.D.M. Gómez-Ramos, A. Pérez-Parada, J.F. García-Reyes, A.R. Fernández-Alba, A. Agüera Use of an accurate-mass database for the systematic identification of transformation products of organic contaminants in wastewater effluents *J. Chromatogr. A*, 1218 (2011), pp. 8002–8012
- [116] M.J. Martínez Bueno, M.M. Ulaszewska, M.J. Gómez, M.D. Hernando, A.R. Fernández-Alba Simultaneous measurement in mass and mass/mass mode for accurate qualitative and quantitative screening analysis of pharmaceuticals in river water *J. Chromatogr. A*, 1256 (2012), pp. 80–88
- [117] J. Nurmi, J. Pellinen, A.-L. Rantalainen Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry *J. Mass Spectrom*, 47 (2012), pp. 303–312
- [118] A.C. Hogenboom, J.A. van Leerdam, P. de Voogt Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography–hybrid linear ion trap Orbitrap mass spectrometry *J. Chromatogr. A* (2009), pp. 510–519
- [119] E.L. Schymanski, H.P. Singer, P. Longrée, M. Loos, M. Ruff, M.A. Stravs, et al. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry *Environ. Sci. Technol*, 48 (2014), pp. 1811–1818
- [120] J. Nurmi, J. Pellinen Multiresidue method for the analysis of emerging contaminants in wastewater by ultra performance liquid chromatography-time-of-flight mass spectrometry *J. Chromatogr. A*, 1218 (2011), pp. 6712–6719
- [121] R. Díaz, M. Ibáñez, J.V. Sancho, F. Hernández Qualitative validation of a liquid chromatography-quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters *J. Chromatogr. A*, 1276 (2013), pp. 47–57

[122] R. Bagnati, E. Davoli Analytical methods for the detection of illicit drugs in wastewaters and surface waters S. Castiglioni, E. Zuccato, R. Fanelli (Eds.), *Illicit Drugs in the Environment: Occurrence, Analysis and Fate Using Mass Spectrometry*, John Wiley & Sons, Inc (2011), pp. 55–67

[123] M. van der Aa, L. Bijlsma, E. Emke, E. Dijkman, A.L.N. van Nuijs, B. van de Ven, et al. Risk assessment for drugs of abuse in the Dutch watercycle *Water Res*, 47 (2013), pp. 1848–1857

1.4 Marie Curie SEWPROF ITN

This work has been made as part of the European Union funded Marie Curie International Training Network (ITN) SEWPROF: A new paradigm in drug use and human health risk assessment: Sewage profiling at the community level. This ITN is multi-disciplinary and multi-institutional, covering subjects from analytical chemistry to statistics and epidemiology and includes eleven institutions in nine countries (Norway, Denmark, Germany, United Kingdom, The Netherlands, Belgium, Switzerland, Italy and Spain). Eleven Early Stage Researchers (ESRs) and four Experienced Researchers (ERs) from all over the world were recruited for this project. The main objective of this project is to “advance knowledge of the epidemiology of (illicit) drug use and to bridge gaps in the available expertise with the ultimate goal of applying this cutting edge interdisciplinary approach within epidemiological studies of societal health” (<http://sewprof-itn.eu/>).

The interdisciplinary nature of this project is exemplified by the institutes and departments involved: from toxicologists (Toxicological Centre, University of Antwerp and Department of Experimental and Clinical Toxicology, University of Saarland) to water and environmental research institutes (KWR Watercycle Research Institute, Department of Environmental Health Sciences (Mario Mario Negri Institute of Pharmacological Research), Ecotoxicology and Risk Assessment (Norwegian Institute for Water Research and Research Institute of Pesticides and Water (University Jaume I)), engineers (Department of Environmental Engineering (Technical University of Denmark) and Urban Water Management (Swiss Federal Institute of Aquatic Science and Technology)), epidemiologists (Centre for Drug and Addiction Research (University of Oslo), drug identification agencies (TICTAC Communications Ltd.) and pure chemists (Department of Chemistry (University of Bath)).

As implied from the title “International Training Network”, research stays for training and research purposes are mandatory. Research stays were thus performed by the PhD candidate at Mario Negri Institute of Pharmacological Research, Milan, Italy; KWR Watercycle Research Institute, Nieuwegein, The Netherlands; University of Saarland, Homburg, Germany; and University of Antwerp, Belgium.

Research stays and brief statement of work undertaken:

Mario Negri Institute of Pharmacological Research: Suspect screening of wastewater and surface water samples from Spain and Italy using an LTQ-Orbitrap, under supervision of Dr. Sara Castiglioni. Duration: 28 April – 26 May 2014. This work resulted in **Scientific Article 3**.

KWR Watercycle Research Institute: Suspect and non-target screening of wastewater samples following injection in LTQ-Orbitrap using advanced deconvolution software, under supervision of Erik Emke and Prof. Dr. Pim de Voogt. Duration: 17 August – 13 September 2014. This work resulted in **Scientific Article 4**.

University of Saarland: *in vitro* metabolism studies of 3-fluorophenmetrazine, under supervision of Prof. Dr. Hans Maurer. Duration: 1-18 June 2015. This work resulted in a co-authored paper (M. Mardal, B. Miserez, R. Bade, T. Portolés, M. Bischoff, F. Hernández, M. R. Meyer; “3-Fluorophenmetrazine, a fluorinated analogue of phenmetrazine: studies on *in vivo* metabolism in rat and human, *in vitro* metabolism in human CYP isoenzymes and microbial biotransformation in *Pseudomonas Putida* and wastewater using GC and LC coupled to (HR)-MS techniques”, SUBMITTED)

University of Antwerp: retention time prediction using artificial neural networks of three different LC columns: biphenyl, C8 and HILIC, under supervision of Dr. Alexander van Nuijs and Dr. Adrian Covaci. Duration: 20 July – 14 August 2015. While this work did provide some good results, further research is required before a publication can be anticipated (see Future Perspectives).

1.5 Scientific Articles Included in this Thesis

The core of this doctoral thesis consists of seven scientific articles published in (or submitted to) peer-reviewed journals within the first quartile of the ranking by Impact factor (ISI web of knowledge). It is worth mentioning that five of these articles were made in collaboration with some internationally-regarded scientists as part of the mandatory research stays mentioned in Section 1.4 as well as separate international collaborations. The role of the PhD candidate within these articles is outlined below.

Scientific Article 2 was made in collaboration with researchers at the Universidad Nacional de Colombia and Universidad Antonio Nariño in Bogotá, Colombia. The PhD candidate was involved with the analysis of these samples in the UHPLC-QTOF as well as the data processing and writing of the article.

Scientific Article 3 was the result of a collaboration with the Mario Negri Institute of Pharmacological Research (MN), Milan, Italy as part of a research stay within the SEWPROF project. Sample collection, preparation and analysis was all performed by the PhD candidate at the University Jaume I (using UHPLC-QTOF) while the candidate analysed the samples at MN following injection in LC-Orbitrap. Furthermore, the candidate drafted and revised the manuscripts. This work was performed independently under the supervision of Prof. Dr. Félix Hernández and Dr. Sara Castiglioni.

Scientific Article 4 resulted from a collaboration with KWR Watercycle Research Institute, Nieuwegein, The Netherlands as part of two separate research stays (the first as a SEWPROF secondment and the second as a Short-Term Scientific Mission funded by COST ES1307). The experimental work (discounting sample preparation) and data analysis was entirely performed at KWR, independently under the supervision of Prof. Dr. Félix Hernández, Prof. Dr. Pim de Voogt and Erik Emke. The candidate drafted and revised the manuscripts.

Scientific Article 6 involved a collaboration with Dr. Leon Barron and Thomas Miller of King's College, London. This work was a follow-up to **Scientific Article 5**, with a subset of 544 compounds used in the development of a retention time prediction method using Artificial Neural Networks. The PhD candidate was involved in the initial development of the molecular descriptor selection and analysis and subsequent implementation for suspect analysis as well as writing and revision of the article.

Scientific Article 7 was the results of a final collaboration between eight institutions within the SEWPROF network. Samples from collected by all institutes and were used investigated using the optimised and validated method for new psychoactive substances, which was performed solely at University Jaume I. The candidate performed all experimental work and drafted the initial manuscript, which was later revised by all co-authors. This work was performed independently under the supervision of Prof. Dr. Juan V. Sancho, Dr. Lubertus Bijlsma and Prof. Dr. Felix Hernandez.

CHAPTER 2

QUALITATIVE SCREENING WITH LIQUID CHROMATOGRAPHY COUPLED WITH HIGH RESOLUTION MASS SPECTROMETRY

2.1 Introduction

Liquid Chromatography coupled to High Resolution Mass Spectroscopy (LC-HRMS) is the technique of choice for the quick detection and (possible) identification of a large number of emerging contaminants in the aquatic environment. HRMS instruments combine full spectrum MS data with high mass resolution which makes them able to accurately measure the mass of any ionisable analyte in a sample. Theoretically, this means that there is a limitless number of compounds able to be detected using HRMS, given that they fit the requirements of the sample treatment, chromatographic treatment and MS ionisation. With the increased mass accuracy and resolution associated with HRMS (up to 100,000) compared to QqQ instruments, compounds of a similar exact mass are more easily differentiated and therefore it is very useful in the derivation of molecular formulae. Isobaric compounds, inseparable based on nominal mass (one decimal place) but containing different molecular formulae and thus exact mass (four decimal places), can hereby be distinguished using accurate mass measurements.

In low resolution instruments, such as QqQ, isobaric compounds can be discriminated only by product ions and retention time. However, when bi-isobaric compounds co-elute or full chromatographic separation is not made, a false positive or negative may be reported. In HRMS, isobaric interferences can be discriminated by reducing the extracted ion chromatogram (XIC) window (down to 0.01-0.05 Da), which increases the signal-to-noise ratio by simultaneously decreasing the chemical noise, therefore improving the reliability and sensitivity of the data. (Hernández et al., 2005; Nurmi and Pellinen, 2011)

The high quality, sensitive and accurate-mass data obtained by HRMS allows the determination of the elemental composition of a compound with high confidence. Because of the resolution, very similar masses can be differentiated by their different elemental compositions. For example, purine ($C_5H_4N_4$) and acetophenone (C_8H_8O) both have the same nominal mass of 120.1, and with the nominal mass resolution of a QqQ, these ions would be indistinguishable. However, based on their different elemental compositions, they have different exact masses (120.0436 and 120.0575 respectively), which could be easily distinguished using an instrument with a higher resolving power, such as a QTOF or Orbitrap. Furthermore, the characteristic isotopic patterns of carbon, bromine, sulphur and chlorine can be utilised to reduce the number of possibilities and help find the correct elemental composition.

All of these characteristics combined make HRMS a suitable instrument for environmental samples analysis, such as surface and wastewater, where there are potentially thousands of compounds warranting investigation. On the one hand, environmental analytical chemists can apply target approaches, being limited when choosing which compounds to focus on, as reference standards, required for quantitative analysis, can be very costly. With a reliance on sensitivity for these analyses, HRMS is not commonly the choice, although it has been used (Bijlsma et al., 2013; Gonzalez-Mariño et al., 2012; Leendert et al., 2015), rather, triple quadrupole (QqQ) instruments (Farré et al., 2012). But on the other hand, qualitative screening methods, making use of databases containing thousands of (known) emerging contaminants and metabolites can be used to get more of a general idea on which compounds are present. This can provide a quick overview of compounds present in the samples. However, one disadvantage associated with this approach is that even with a database containing thousands of compounds, there may still be peaks present that are unaccounted for. Alternatively, analyses could be made without reference standards or databases, but based solely on the data obtained from the LC-HRMS. These three ideas are, indeed, the three strategies used for the detection and identification of emerging contaminants in environmental analyses: target (list of targeted compounds to be searched, with reference standards available), suspect (list of targeted compounds to be searched, without reference standards available to the laboratory), and non-target (discovering of unknowns, without any selection of targets to be searched).

One disadvantage of target and suspect screening is the “Matthew Effect” as noted by Daughton (Daughton, 2014), which basically states that compounds that have been detected/identified in previous studies are more likely to be included in target/suspect screening strategies. Although this seemingly self-fulfilling selection bias could be construed as chemical or environmental ignorance, there are several plausible reasons why not all compounds are able to be studied:

- Newly registered/developed pharmaceuticals/drugs of abuse may gain rapid market penetration and ascend the rankings of prescription data and consumer consumption faster than targeted analysis are able to be developed.
- Some compounds are not easily applicable for analytical methods, with difficulties resulting from their physicochemical characteristics, from polarity to stability. Therefore, some compounds which may be present will not (or cannot) be demonstrated.
- Some compounds are simply ignored, due to wrongly predicted or performed environmental prevalence studies. Alternatively, even correctly performed predictions or studies can be

confounded due to the extreme (and irreproducible) conditions associated with the transport from the source.

The first of these is particularly relevant for new psychoactive substances (NPS; more in Chapter 4), which are readily produced and changed depending on consumers and legal regulations, with analytical methods often one step behind the production and user uptake. Furthermore, with costly reference standards necessary for target (and ultimately, suspect) analyses, most laboratories are (rightly) apprehensive when it comes to purchasing new reference standards for compounds which may not be present in the environment or for which limited data is available.

The second reason is one which is very important, often overlooked by environmental epidemiologists and non-chemists. It is impossible to make a single analytical method for all compounds. Even making a small, targeted method for tens of compounds is time-consuming. Conversely, in wide-scope suspect screening methods, a compromise is made so as to ensure that as many compounds as possible can be detected. This can mean that very polar and non-polar compounds can sometimes be overlooked.

The last reason is difficult to foresee. It is impossible on a lab-scale to reproduce the conditions seen in the environment, in terms of sunlight, water, possible sorption and interaction with other compounds. Even with studies performed well, these unknown interactions could result in compounds initially not expected to be present, to be present.

This bias and “Matthew Effect” can be overcome by using non-target analysis, where no prior information is known at all, so no selection or availability bias can be inputted. However, the compounds detected are limited by the analytical methodology, as explained above, where some compounds may not be able to be detected due to their particular chemical characteristics. Moreover, it is unrealistic to investigate the identity of every peak appearing in a total ion chromatogram (TIC), with the additional limitation that “relevant compounds” from the environmental or WBE point of views, are commonly present at very low concentration in complex unknown matrices, such as wastewater samples.

It is difficult to assign one HRMS instrument (i.e. TOF or Orbitrap) as being superior, as their use depends on the aim of a particular study. Most recent publications include TOF MS instruments with mass accuracies typically less than 5ppm and mass resolving power of 20,000 - 40,000 FWHM depending on the instrument. The scan speed attainable in recent TOF instruments falls in the range of 10-100 scans/s, more than enough for UHPLC separations, leading to shorter run times, enabling

sufficient chromatographic resolution to minimize co-elution of compounds with close m/z values. One further advantage of a (Q)TOF is its ability to perform simultaneous full-scan and MS/MS (Martínez Bueno et al., 2012) (or “pseudo”-MS/MS (Hernández et al., 2011)), meaning that information pertaining to fragment ions can be collected in a single injection, rather than having to reinject a second time. The aforementioned “pseudo”-MS/MS mode, MS^E , of the QTOF was used in both **Scientific Article 2** and **3** for suspect screening. The ability to acquire low-energy (LE) and high-energy (HE) data at the same time, without the need to preselect a precursor molecule, facilitated data analysis. By applying LE in the collision cell, fragmentation is minimized, and the information obtained corresponds normally to non-fragmented ions, related to the parent molecule. However, at HE, fragmentation will take place, resulting in abundant fragments (Hernández et al., 2011; Plumb et al., 2006). The MS^E approach is also known as All-Ion Fragmentation or broadband collision-induced dissociation, depending on the manufacturer.

Orbitrap has a higher resolution than TOF (up to 150,000 FWHM), which gives obvious advantages in the analysis, particularly in samples of high complexity, and is considered better for quantitative analysis, as it affords a linear dynamic range of up to three-four orders of magnitude, at least one order of magnitude higher than the TOF instruments (Agüera et al., 2013; Krauss et al., 2010). The main drawback of this instrument is its scanning speed, which is inverse of mass resolution. For example, at a mass resolution of 100,000, a scan speed of only 1 scan/s is attainable. It can, however, be coupled to UHPLC, but only when using a sufficiently low mass resolution (i.e. 10,000 FWHM for a scan speed of 10 scans/s).

Sample treatment associated with wide-scope screening methods is an important aspect, and can be a limitation when dealing with many compounds of quite different physicochemical characteristics. A simple sample treatment can be applied, such as SPE with Oasis HLB for pre-concentration and to enable detecting the analytes at the low concentration levels normally present, which can be thought of as “universal” and generic (Hernández et al., 2015), thereby allowing as many compounds as possible being detected following sample analysis. Alternatively, to gain more specific information on the compounds present multiple SPE cartridges could be used (Kern et al., 2009; Schymanski et al., 2014), or, if, in advance, a screening was made on a more specific set of compounds, different sample treatment can be applied – such as adding activated carbon for more polar compounds (Schollée et al., 2015).

This chapter focusses on the use of both UHPLC-QTOF-MS and LC-LTQ-Orbitrap-MS in the detection and identification of emerging contaminants in environmental waters.

2.2 Scientific Articles

This chapter includes three scientific articles, “*LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia*” (**Scientific Article 2**), “*Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS*” (**Scientific Article 3**) and “*Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools*” (**Scientific Article 4**). They concern suspect and non-target screening for the identification of pharmaceuticals, illicit drugs and their metabolites in environmental waters of different geographic and aquatic origin. All of the studies incorporated the same sample treatment: with samples being filtered through a membrane filter (0.45µm) to remove any solid particulates, then extracted on HLB SPE cartridges. This generic sample treatment was employed due to the wide-scope nature of the studies. With no specific compounds under investigation, no specific sample treatment could be applied.

Scientific Article 2 was envisaged as a very comprehensive initial study into the pharmaceuticals and illicit drugs present in waste and surface waters of Bogotá, Colombia. This was seen as a first step towards target analysis of the detected compounds, to get a more sensitive and quantitative idea of their existence. A database of more than 1000 compounds was used in the detection/identification process. After a simple SPE extraction with a generic polymeric cartridge as Oasis HLB, sample extracts were injected in MS^E mode in the UHPLC-QTOF so as to gain information relating to parent and fragment ions in a single injection. In the end, 30 parent compounds and metabolites could be confirmed, based on the presence of at several ions at accurate mass, while a further five could not be confirmed as there was only one ion. Six compounds, were tentatively identified based on fragment ions, three of which were tentatively identified based on a common fragmentation pathway.

A similar procedure was followed in **Scientific Article 3**, but employing a smaller list of compounds, those, which, based on previous experience, would be more likely to be found in environmental waters. To take this suspect analysis a step further, it was also performed on an LTQ-Orbitrap instrument. Because the methods and instruments were completely different, a comparison was not made, rather to see whether any additional compounds could be detected by employing a different instrument. After injection, the data was analysed using the respective vendors’ software. When one compound was found using one combination of instrument and software but not the other, the raw

data was investigated. In the end, little difference was seen between the instruments, proving that they are both powerful instruments, useful in the qualitative identification of emerging contaminants.

Scientific Article 4 investigated a non-target approach for the tentative identification of pharmaceuticals and illicit drugs. Following injection in an LC-LTQ-Orbitrap, two software types were utilised for hidden target screening (MsXelerator and Sieve 2.1). Two databases were used, one similar to that employed in **Scientific Article 3** as well as ChemSpider, to detect possible compounds following automated exact mass extraction. Combining the detected compounds from the two hidden target methods, a target list was made for subsequent MS/MS investigation, in order to tentatively identify compounds based on product ions and their comparability with literature and/or the online mass spectral database MassBank.

The results of these three articles have been published as follows:

LC-QTOF MS screening of more than 1000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia *F. Hernández, M. Ibáñez, A. M. Botero-Coy, R. Bade, M. Bustos-Lopez, J. Rincon, A. Moncayo, L. Bijlsma*, *Analytical and Bioanalytical Chemistry* 407, **2015**, 6405-6416.

Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS *R. Bade, N. I. Rousis, L. Bijlsma, E. Gracia-Lor, S. Castiglioni, J. V. Sancho, F. Hernández*, *Analytical and Bioanalytical Chemistry* 407, **2015**, 8979-8988.

Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools *R. Bade, A. Causanilles, E. Emke, L. Bijlsma, J. V. Sancho, F. Hernandez, P. de Voogt* *Science of the Total Environment* UNDER REVISION.

2.2.1 Scientific Article 2

**LC-QTOF MS SCREENING OF MORE THAN 1,000
LICIT AND ILLICIT DRUGS AND THEIR
METABOLITES IN WASTEWATER AND SURFACE
WATERS FROM THE AREA OF BOGOTÁ, COLOMBIA**

Félix Hernández, María Ibáñez, Ana-Maria Botero-Coy, Richard Bade, Martha
Cristina Bustos-López, Javier Rincón, Alejandro Moncayo, Lubertus Bijlsma

Analytical and Bioanalytical Chemistry (2015) 407: 6405-6416

LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia

Félix Hernández¹ · María Ibáñez¹ ·
Ana-María Botero-Coy¹ · Richard Bade¹ ·
Martha Cristina Bustos-López² · Javier Rincón³ ·
Alejandro Moncayo³ · Lubertus Bijlsma¹

Received: 25 February 2015 / Revised: 12 May 2015 / Accepted: 20 May 2015 / Published online: 18 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Topical Collection: High-Resolution Mass Spectrometry in Food and Environmental Analysis

Abstract

A large screening of around 1,000 emerging contaminants, focused on licit and illicit drugs and their metabolites, has been made in urban wastewaters (both influent and effluent) and surface waters from the area of Bogotá, Colombia. After a simple generic solid-phase extraction (SPE) step with Oasis hydrophilic-lipophilic balanced (HLB) cartridges, analyses were made by ultra high-performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (UHPLC-QTOF MS) under MS^E mode (sequential acquisition of mass spectra at low energy (LE) and high collision energy (HE)). Accurate mass measurements and the information provided by MS^E on the presence of the (de)protonated molecule and fragment ions allowed the reliable identification of the compounds detected, even without reference standards being available in some cases (tentative identification). The compounds most frequently found were acetaminophen/paracetamol, carbamazepine and its dihydro-dihydroxylated metabolite, clarithromycin, diclofenac, ibuprofen, gemfibrozil, lincomycin, losartan, valsartan, the two metabolites of metamizole (4-acetamido-antipyrine and 4-formylamino-antipyrine), sucralose, and cocaine and its main metabolite benzoylecgonine. Caffeine, the sweetener saccharin, and two hydroxylated metabolites of losartan were tentatively identified in almost all samples analyzed. Pharmaceutical lidocaine was tentatively identified and subsequently confirmed with reference standard. For the first time, a general overview of the occurrence of drugs and their metabolites in the aquatic environment of Colombia has been reported. In the near future, target methodologies, typically based on liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), will need to be set up for accurate and sensitive quantification of the contaminants selected on the basis on the information provided in the present paper.

Keywords

Screening, ultra high-performance liquid chromatography, time-of-flight mass spectrometry, drugs, water, Colombia

Introduction

Changes in industry, agriculture, and urban development present interlinked challenges to water quality [1, 2]. Due to the increasing use of pharmaceuticals and personal care products (PCPs) as well as veterinary drugs and illicit drugs, the amount of these newly emerging chemicals detected in waters raises environmental and health concerns. Many of these contaminants survive the passage through conventional wastewater treatment processes resulting in the growing discharge of these compounds into receiving waters where their presence is increasingly common [3, 4]. However, little information is available from Latin America despite the fact that this region undergoes rapid land, economical, and social changes [5] and where, in some areas, the discharge of raw sewage into rivers, lakes, and reservoirs is widely practiced [6–8]. Concern over the presence of these contaminants is well founded as these rivers not only flow through tropical areas rich in biodiversity, but also surface waters are also used as a source for human consumption after conventional chlorination processes or directly as irrigation water without any treatment.

Monitoring licit and illicit drugs is generally based on multi-residue methodologies using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for target substances considered as harmful for human health or the ecosystem. Analyses are commonly focused on a limited number of priority compounds, and the main objective is the accurate quantification of the target analytes. However, the use of pharmaceuticals and veterinary drugs differs temporally and spatially between regions and countries due to different marketing, regulations, prescription practices, etc. [9]. Thus, drugs used in Latin America may differ from those generally prescribed or consumed in Europe and North America. Accordingly, target approaches applied for routine analysis of water samples in one country might ignore important contaminants from another country, as they are simply not included in the scope of the method. Therefore, a large screening of contaminants is of high interest in order to define priority compounds and to subsequently set up target methodologies for monitoring. To this end, high-resolution mass spectrometry (HRMS) plays an important role for screening of emerging contaminants, in relevant fields like environmental pollution [10, 11].

Hybrid HRMS, such as a quadrupole time-of-flight mass spectrometry (QTOF MS) in MS^E or all ion MS/MS mode (the term depends on the manufacturer), has demonstrated its use as an advanced tool that allows the investigation of hundreds of compounds in the same run, making use of homemade databases [12–14]. The presence of compounds in samples can be investigated once the analysis has been performed and data acquired, without being dependent on the pre-selection of analytes. It allows detection/rapid screening for compounds by means of searching the exact mass of the (de)protonated molecules in the full spectra generated using low collision energy (LE function). Confirmation with reference standards, or tentative identification, could subsequently be performed searching for exact masses of fragment ions in the acquired spectra at high collision energy (HE), where fragmentation of the molecule is promoted.

We performed a previous monitoring in the irrigate district of Usosaldaña, an important agricultural area in Colombia, mainly devoted to the cultivation of rice. Analysis of surface water and soil samples by TOF MS, coupled to both gas and liquid chromatography revealed the presence of several pesticides and metabolites in most of the samples. Also, some personal care products were identified [15]. However, there is little or no information on the presence of emerging organic contaminants in the area of Bogotá. In the present work, a large screening of pharmaceuticals belonging to different therapeutic groups, veterinary drugs, X-ray agents, PCPs (preservatives and UV filters), sweeteners, illicit drugs, and a notable number of metabolites has been performed in urban wastewaters (both influent and effluent) and receiving surface waters from the surrounding area of Bogotá. A customized database with more than 1,000 emerging contaminants was developed including pharmaceuticals frequently prescribed in Colombia. As far as we know, this is the first wide-scope screening based on HRMS that has been applied to this aim. Henceforth, future monitoring in this area can be focused on the compounds identified in this initial step.

Material and methods

Chemicals and standards

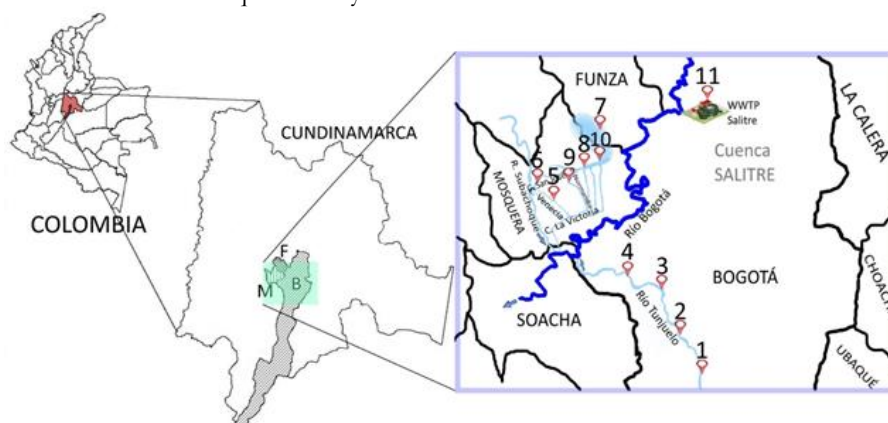
Eight hundred and sixty-five human and veterinary pharmaceuticals including metabolites (see Electronic Supplementary Material (ESM), Table S1), 29 X-ray agents (ESM Table S2), 20 UV filters (ESM Table S3), 4 preservatives (ESM Table S4), 9 sweeteners (ESM Table S5), and 130 illicit drugs including metabolites (ESM Table S6) were studied, of which 215 reference standards were available (asterisk in ESM Tables S1–S6).

Reference standards of these compounds were purchased from Across Organics (Geel, Belgium), Aventis Pharma (Madrid, Spain), Bayer Hispania (Barcelona, Spain), Cerilliant (Round Rock, TX, USA), Fluka (Buchs, Switzerland), Dr. Ehrenstorfer (Augsburg, Germany), Fort Dodge Veterinaria (Gerona, Spain), LGC Promochem (London, UK), National Measurement Institute (Pymble, Australia), Riedel-de Haën (Seelze, Germany), Sigma-Aldrich (St. Louis, MO, USA), Toronto Research Chemicals (Ontario, Canada), Vetoquinol Industrial (Madrid, Spain), and Witega (Berlin, Germany). All reference materials had purities higher than 98 % (*w/w*), except for marbofloxacin and pefloxacin, which had purities higher than 93 %.

HPLC-grade methanol (MeOH), HPLC-grade acetonitrile (ACN), sodium hydroxide (NaOH) (>99 % *w/w*), and formic acid (HCOOH) (>98 % *w/w*) were purchased from Scharlau (Barcelona, Spain). Leucine enkephalin was purchased from Sigma-Aldrich (Madrid, Spain). HPLC-grade water was obtained by purifying demineralized water in a Milli-Q plus system from Millipore (Bedford, MA, USA). Solid-phase extraction (SPE) cartridges used were 3-cm³ Oasis hydrophilic-lipophilic balanced (HLB) (60 mg) from Waters (Milford, MA, USA).

Samples

Seven influent wastewater (IWW) samples and seven effluent wastewater (EWW) samples were taken from the Salitre wastewater treatment plant (WWTP), in Northwest Bogotá D.C., Colombia, which serves a population of approximately three million inhabitants. The 24-h composite samples were collected over seven consecutive days in March 2014 (starting on Wednesday, March 12, and ending on Tuesday, March 18) in high-density polystyrene bottles, immediately centrifuged, and stored in the dark at $-20\text{ }^{\circ}\text{C}$ until analysis. In addition, ten grab surface water (SW) samples were collected from around Bogota (Fig. 1) in July 2013, concerning areas of interest along the Tunjuelo River (samples 1–4) and in La Ramada Irrigation District (samples 5–10). Seven individual samples (250 mL) were taken for each sampling point. These were then combined to form a composite sample to provide a more complete overview of each area. An aliquot was then taken for the sample treatment and subsequent analysis.



Sampling points

Number	Name/Description
1	Landfill, Doña Juana (Tunjuelo river)
2	San Benito (tanneries area) (Tunjuelo river)
3	Guadalupe (slaughterhouse) (Tunjuelo river)
4	Bosa (Tunjuelo river)
5	Irrigation channel San Jose- La Victoria (Ramada irrigation area)
6	Irrigation channel San Jose- Los Pinos (Ramada irrigation area)
7	Wetland Güali-Tres Esquinas (Ramada irrigation area)
8	Canal C. Agricultural Center, Marengo (Ramada irrigation area)
9	Canal C. Agricultural Center, Marengo (Ramada irrigation area)
10	Exit of swamp (Ramada irrigation area)
11	Wastewater, Salitre WWTP Seven influent and seven effluent wastewater samples (one whole week)

Fig. 1 Location of sampling sites

Sample treatment

An SPE step was applied prior to analysis to pre-concentrate the sample. All samples were filtered through 0.45- μm mixed cellulose ester membrane filter (Whatman, Dassel, Germany). SPE was performed using Oasis HLB cartridges (60 mg). The cartridges were conditioned by washing and rinsing with 6 mL MeOH and 6 mL Milli-Q water. The water samples (IWW was four times diluted with Milli-Q water, i.e., 25 mL sample in 100 mL; EWW and SW were 100 mL, no dilution) were loaded onto the cartridges, percolated by gravity (flow rate around 3 mL/min), and vacuum dried for approximately 15 min. Analytes were eluted with 5 mL MeOH. The extracts were evaporated to dryness at 35 °C under a gentle stream of nitrogen and reconstructed in 1 mL of 10:90 MeOH/H₂O. Analyses were performed by injecting 20 μL of the final extract into the UHPLC-QTOF MS.

Instrumentation

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (XEVO G2 QTOF, Waters Micromass, Manchester, UK), using a Z-Spray electrospray ionization (ESI) interface operating in both positive and negative ion mode. The chromatographic separation was performed using an Acquity UPLC BEH C18 1.7- μm particle size column 100 \times 2.1 mm (Waters) at a flow rate of 300 $\mu\text{L}/\text{min}$. The mobile phases used were A = H₂O with 0.01 % HCOOH and B = MeOH with 0.01 % HCOOH. The initial percentage of B was 10 %, which was linearly increased to 90 % in 14 min, followed by a 2-min isocratic period and, then, returned to initial conditions during 2 min. The total run time was 18 min. Nitrogen was used as drying gas and nebulizing gas. The desolvation gas flow was set at 1,000 L/h and the cone gas at 80 L/h. TOF MS resolution was approximately 20,000 at full width at half maximum (FWHM) at m/z 556.

MS data were acquired in centroid mode over an m/z range of 50–1,000 Da. Data were acquired in both positive and negative ionization mode in two separate runs. A capillary voltage of 0.7 and 2.5 kV was used in positive and negative ionizations mode, respectively. A cone voltage of 20 V was used. Collision gas was argon 99.995 % (Praxair, Valencia, Spain). The desolvation temperature was set to 600 °C and the source temperature to 130 °C. The column temperature was set to 40 °C.

For MS^E experiments, two acquisition functions with different collision energies were created: the LE function, selecting a collision energy of 4 eV, and the high energy (HE) function, with a collision energy ramp ranging from 15 to 40 eV in order to obtain a greater range of fragment ions. The scan time for both LE and HE function was 0.4 s.

For elucidation of losartan metabolites, MS/MS experiments at different collision energies (10, 20, 30, 40, and 50 eV) were performed to promote higher fragmentation and also to ensure that fragment ions observed in MS^E come from the parent ion.

Data processing

Processing of MS data was done using ChromaLynx XS application manager (within MassLynx v 4.1; Waters Corporation). The following parameters were used for screening: mass window 150 ppm (for positive ID ≤ 5 ppm), isotope fit as well as retention time (maximum deviation of ≤ 2.5 %) and fragmentation, when available. Software-specific settings were as follows: peak width at 5 % height, 6 s; peak-to-peak baseline noise, 1,000; and threshold absolute area, 200.

Results and discussion

Detection, confirmation and tentative identification

The full-spectrum accurate mass data, generated by QTOF under MS^E mode, were inspected in ChromaLynx XS (within MassLynx) using a homemade database containing 1,057 LC-amenable organic contaminants (see ESM, Tables S1–S6), which is continuously being updated. Analytes were mostly selected based on our own experience and on existing compound lists encountered in the literature on LC-MS methods for determination of organic contaminants. Empirical data (retention time, adduct information, and/or fragment ions) obtained from compounds of which reference standards (up to 215) were available in our laboratory was included. This information was used for an easier and more straightforward detection and unambiguous confirmation of the identity of the compound. Furthermore, pharmaceuticals frequently prescribed in Colombia were also included. For those compounds for which reference standards were not available at our laboratory, the only information included in the database was the name and elemental composition of the compounds. Here, the tentative identification was a manual and more time-consuming process, incorporating Internet resources, such as MassBank [16] or MetLin [17], to look for information about possible fragment ions.

The confidence in HRMS-based identification is an important issue as pointed out by Schymanski et al. [18]. Discussion on how to best communicate confidence regarding identification for the exchange of results via literature and databases is ongoing, but criteria are needed and should be well reported. Criteria used in this study were based on the availability (detection and confirmation of the identity) or non-availability (tentative identification) of reference standards. The information given by QTOF MS spectra, both LE and HE accurate mass spectra, and retention times allowed the compounds to be distinguished using different possibilities:

- *Detection*, based on the presence of one accurate mass ion (mass error ≤ 5 ppm) and retention time agreement (maximum deviation ≤ 2.5 %)
- *Confirmation of the identity*, with at least two accurate mass ions (≤ 5 ppm) and retention time (≤ 2.5 %)
- *Tentative identification*, with at least two accurate mass ions justified by literature data and/or compatibility with candidate chemical structure

For those compounds containing chlorine or bromine atoms, the characteristic isotopic profile should be observed. However, we considered the isotopic pattern as a prerequisite for identification, but not as valuable as an accurate mass fragment ion. Therefore, when only those ions corresponding to the parent compound and the isotopic peak were present, we considered the compound as detected, not as confirmed.

In the case of tentative identification, reference standards and additional MS/MS experiments would be necessary for final unequivocal confirmation of the compounds. However, we do not need expensive reference standards for all 1,057 compounds in the database, only for the positive findings. From our previous experience, the subsequent acquisition of reference standards has allowed confirmation of nearly all tentative identifications. This supports the high degree of reliability of tentative identification by using this technique [19].

Screening of Colombian water samples

In order to perform a large screening, a non-selective sample treatment, and a generic chromatographic separation were chosen to broaden the system applicability to as many compounds as possible. Due to the large level of dilution in these types of samples, SPE was performed, using a generic Oasis HLB cartridge for pre-concentration and to enable detecting the analytes at the low concentration levels normally present. Using this analytical procedure and from the information obtained by UHPLC-QTOF MS under MS^E mode, several compounds could be detected, confirmed, and/or tentatively identified in all type of Colombian water samples analyzed (Table 1).

Samples were screened for pharmaceuticals belonging to different therapeutic groups, veterinary drugs, X-ray agents, PCPs (preservatives and UV filters), sweeteners, illicit drugs, and several metabolites. The compounds most frequently detected were the analgesics/anti-inflammatory drugs *N*-acetyl-*p*-aminophenol (commonly known as acetaminophen or paracetamol), diclofenac, ibuprofen, and lidocaine. The antibiotics clarithromycin and lincomycin, the angiotensin II antagonists valsartan and losartan, the anti-epileptic carbamazepine, the beta-blocker metoprolol, the lipid regulator gemfibrozil, and the X-ray agent iopromide were also found. It is interesting to remark the detection or tentative identification of several metabolites such as those of metamizole (4-acetylamino-antipyrine and 4-formylamino-antipyrine), carbamazepine-10,11-dihydro-10,11-dihydroxy, and three metabolites of losartan (losartan carboxylic acid and two hydroxy-losartan isomers). Furthermore, the sweeteners acesulfame, sucralose, saccharin and the psychoactive drugs, cocaine and its main metabolite

benzoylecgonine, as well as caffeine were frequently detected or tentatively identified. UV filters were not detected, while the preservatives methylparaben and propylparaben were found, but only in wastewater samples.

These data are consistent with studies on over-the-counter medicines in Colombia, where analgesics are most frequently sold (60 %), in Bogotá [20]. In addition, 11 % of the population of Bogotá has been diagnosed with hypertension of which 5.5 % take mainly losartan, valsartan, and metoprolol for its control [21] and regarding antibiotics, clarithromycin and lincomycin are most often prescribed [22]. Based on general population surveys, cocaine is the second most consumed illicit drug, but the main illicit psychoactive drug consumed in Colombia is cannabis [23]. However, the main metabolite of cannabis (11-nor-9-carboxy- Δ^9 -tetrahydrocannabinol (THC-COOH)), which is commonly used as biomarker in wastewater for cannabis consumption, was not detected. This was probably due to the low sensitivity of the technique applied for this compound. The analytical problems found in the determination of this compound have been widely recognized in the literature [24–26].

Most compounds were found in both wastewaters (influent and effluent) and surface waters suggesting incomplete removal by WWTPs and possible disposal to receiving surface waters. The pharmaceuticals carbamazepine and its metabolite carbamazepine-10,11-dihydro-10,11-dihydroxy, acetaminophen, diclofenac, ibuprofen, the highly consumed caffeine, cocaine and its metabolite benzoylecgonine are widely occurring contaminants in the aquatic environment, having been reported to be also present in Brazilian surface waters [5, 7].

Table 1. Positive findings by UHPLC-(Q)TOF MS in 10 surface water samples. Results obtained in influent and effluent samples.

COMPOUND	SURFACE WATER (number of the sampling site)										INFLUENT* (n=7)	EFFLUENT* (n=7)	
	1	2	3	4	5	6	7	8	9	10			
<i>Pharmaceuticals and veterinary drugs</i>													
4-Acetamido antipyrine	√	√	√	√	√	√	√	√	√	√	√	√	√
4-Formylamino antipyrine		√	√	√	√	√	√	√	√	√	√	√	√
Acetaminophen/paracetamol	√			√		√	√		√	√	√	√	√
Carbamazepine	√	√	√	√	√	√	√	√	√	√	√	√	√
Carbamazepine 10,11-dihydro-10,11-dihydroxy	√	√	√	√	√	√	√	√	√	√	√	√	√
Clarithromycin		√	√					√	√	√	√	√	√
Clindamycin						√			√	√			
Codeine				•									•
Diclofenac		√	√	√							√	√	√
Dimetridazole		√	√										
Gabapentin													•
Gemfibrozil	√	√	√	√		√	√	√	√	√	√	√	
Ibuprofen		√		√				√			√	√	√
Iopromide				√									√
Irbesartan											√	√	√
Ketoprofen				√							√	√	√
Levamisole			√	√									√
Lidocaine**	⊕√	⊕√	⊕√	⊕√		⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√
Lincomycin		√	√	√		√	√	√	√	√			√
Losartan	√	√	√	√	√	√	√	√	√	√	√	√	√
Losartan, carboxylic acid**	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√	⊕√

Losartan, hydroxy (1)	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Losartan, hydroxy (2)	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Metformin		⊕	⊕	⊕							⊕	⊕
Metoprolol		•	•	•		•		•	•	•	•	•
Metronidazole	√	√	√	√								
Naproxen	√	√	√	√							√	√
Pirantel			⊕	⊕								
Salbutamol				√								
Sulfamethoxazole		√	√	√							√	√
Trimethoprim		√	√	√							√	√
Valsartan		√	√	√		√	√	√	√	√	√	√
<i>Psychoactive drugs</i>												
Benzoylcegonine	√	√	√	√		√	√	√	√	√	√	√
Caffeine	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Cocaine	√	√	√	√		√		√	√	√		√
<i>Preservatives</i>												
Methylparaben												√
Propylparaben											√	√
<i>Sweeteners</i>												
Acesulfame		•	•	•	•	•	√	•	•	√	√	√
Saccharin	⊕			⊕	⊕	⊕	⊕	⊕	⊕		⊕	⊕
Sucralose	•	•	•	√	•	•	√	√	√	√	√	√

Dot (•) detected, not confirmed (one accurate mass ion <5 ppm + retention time <2.5 %), *square root* (√) confirmed with at least two accurate mass ions (<5 ppm) and retention time (<2.5 %) with reference standard, *circled plus* (⊕) tentatively identified (at least two accurate mass ions justified by literature data and/or compatible with the candidate chemical structure) ^aResults given for IWW and EWW correspond to seven samples (one whole week). The compounds indicated as dot, circled plus, or square root were found in at least six out of seven samples analyzed. ^bThese compounds were firstly tentatively identified in the samples and afterwards confirmed with reference standards

In relation to surface waters, the presence of pharmaceuticals was more noticeable in samples collected from the Tunjuelo River (samples one to four)—a tributary of the Bogotá River—in the south of Bogotá. In this area, around 2.5 million inhabitants discharge their sewage directly to this river without any treatment. In addition, solid urban waste are located here (Doña Juana landfill, Fig. 1), and potential leaching surely contributes to the pollution in the area. This can explain that the profile of emerging contaminants found in these surface waters and in wastewaters collected from Salitre WWTP was rather similar, in both cases reflecting the drug consumption of a large population (2.5–3 million inhabitants).

In contrast, the irrigation area of Ramada—supplied by the Bogotá river in its northern part—may be affected by industrial and urban waste, albeit with a lower population (around 600,000 inhabitants). Although these surface water samples (five to ten) had several contaminants in common with those samples from the Tunjuelo River and with wastewater samples, the number of compounds detected was notably lower. Despite this fact, the presence of emerging contaminants may have undesirable consequences for public health as these waters are mainly used for irrigation on this agricultural area. A more in-depth study would be required to quantify the compounds detected in water samples from this area for better knowledge and estimation of potential human risks.

Below, some examples are given in detail, to illustrate the screening strategy applied in this study, which led to the detection, confirmation, or tentative identification of several compounds. The different possibilities considered in this work will be taken into account: presence/absence of reference standards in the laboratory and searching for known and unknown metabolites on the basis of common fragmentation pathway between parent compound and possible metabolites.

1. Reference standards available

Up to 215 reference standards were available in our laboratory and, therefore, experimental data such as retention time and fragment ions could be included in the homemade database. This information was used for easier detection and unambiguous confirmation of the identity of the compounds. Figure 2 shows the unequivocal confirmation in surface water of 4-acetamido-antipyrine, a metabolite of the widely used analgesic metamizole. The LE spectrum (Fig. 2a, top) of the peak at 3.78 min showed the m/z corresponding to the protonated molecule (m/z 246.1237) with a mass error of -2.4 ppm. The HE spectrum (Fig. 2a, bottom) also showed the remaining protonated molecule and five fragment ions with mass errors all below 5 ppm. In this figure, narrow-window eXtracted Ion Chromatograms (nw-XICs) for the six m/z ions are also depicted, with a chromatographic peak at exactly the same retention time.

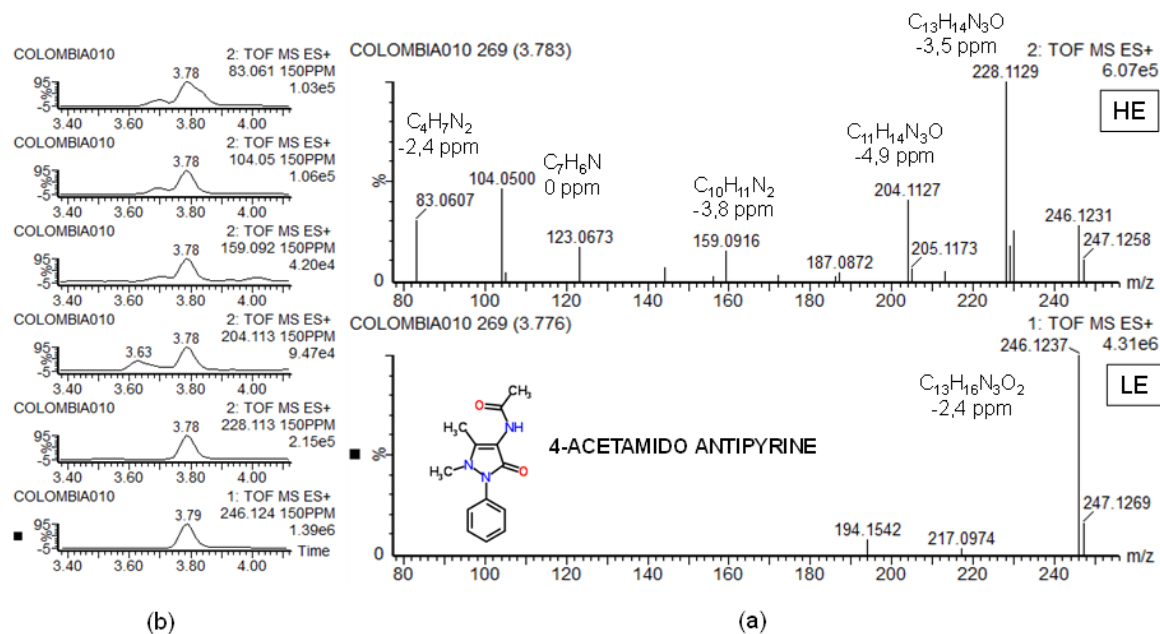


Fig. 2 Detection and identification of metabolite 4-acetylamino-antipyrine in surface water by UHPLC-QTOF MS. (a) Low energy (LE, bottom) and high energy (HE, top) mass spectra for the peak at 3.78 min. (b) eXtracted Ion Chromatograms (XICs) at 150 ppm mass window for $[M+H]^+$ in LE and main fragments in HE

In a similar way, most of the detected compounds could be confirmed in the samples (see the compounds marked as square root in Table 1). The compounds metoprolol and codeine could not be confirmed despite the reference standards were available, as only the protonated molecule was observed at the corresponding retention time. In these cases, the compounds were marked as detected (dot in Table 1). Additional analysis would be required to promote the fragmentation, using higher collision energies in new MS/MS experiments.

2. Reference standards not available

When reference standards were unavailable, the only information included in the database was the name and elemental composition of the contaminants. Even in this situation, a tentative identification could be performed on the basis of the relevant information on the accurate mass of the (de)protonated molecule and the fragment ions offered by QTOF MS^E. Figure 3 illustrates the potential of this screening approach with two case studies: the detection and tentative identification of lidocaine and caffeine in effluent wastewater.

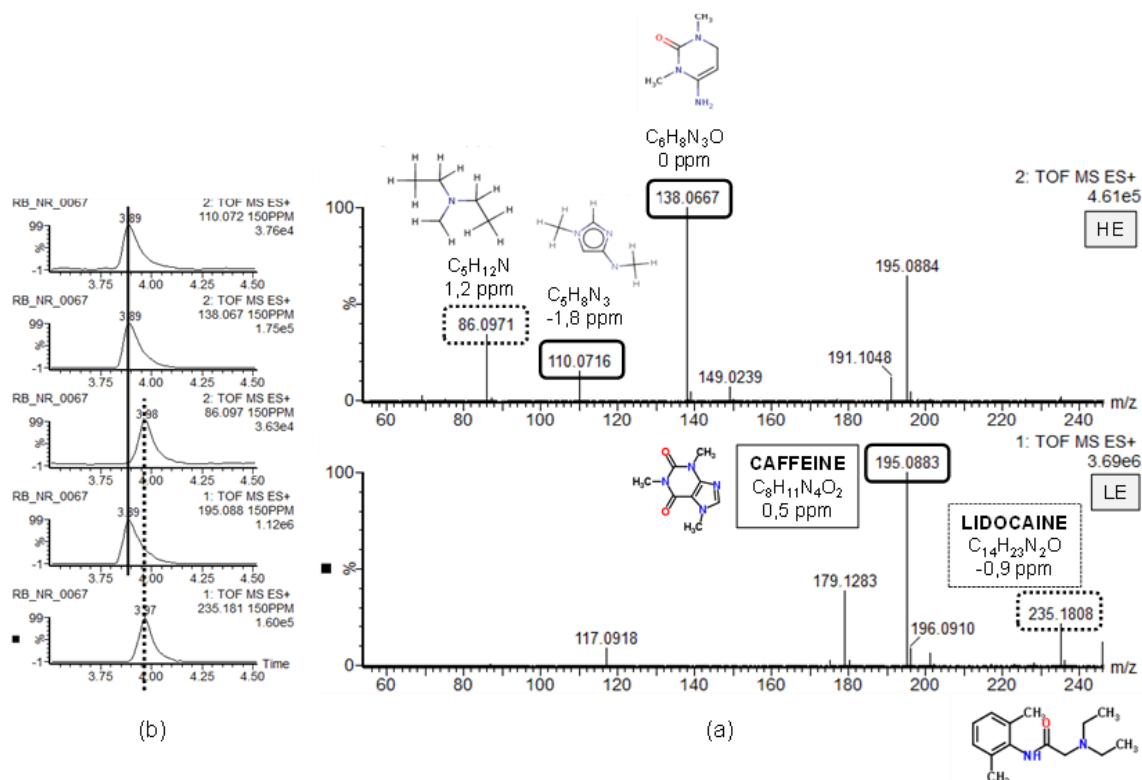


Fig. 3 Tentative identification of lidocaine and caffeine in effluent wastewater by UHPLC-QTOF MS. (a) LE and HE mass spectra for the peaks at 3.89 and 3.98 min in positive ionization mode. (b) XICs at 150 ppm mass window for $[M+H]^+$ of lidocaine (3.98 min) and caffeine (3.89 min) in LE and for the fragment ions (m/z 86, 138, and 110) in HE

The LE spectrum in ESI positive of an abundant chromatographic peak at 3.97 min showed an intense signal at m/z 235.1808 (Fig. 3a, bottom). This might correspond to the protonated molecule of lidocaine (C₁₄H₂₃N₂O, expressed as protonated molecule), with a mass error of -0.9 ppm in relation with the theoretical exact mass. The LE spectrum also showed an important signal at m/z 195.0883 (retention time 3.89 min) which could be attributed to caffeine (C₈H₁₁N₄O₂, 0.5 ppm mass error). The HE spectrum showed three fragment ions at m/z 138.0667 (C₆H₈N₃O), 110.0716 (C₅H₈N₃), and 86.0971 (C₅H₁₂N), all with mass errors below 2 ppm (Fig. 3a, top). At this point, UHPLC was a valuable tool for selecting almost co-eluting fragment ions that might correspond to different precursor ions. Thus, the fragments at m/z 138 and 110 were related to caffeine, whereas the fragment at m/z 86 was related to lidocaine. The structures of the fragment ions were justified on the basis of their measured accurate masses, which was moreover in agreement with the information available in scientific literature [16, 17]. Lidocaine could be tentatively identified and subsequently confirmed after acquiring its reference standard (circled plus and square root in Table 1). The confirmation of caffeine would require the injection of the reference standard, which at the time of writing this paper was not available at our laboratory (we did not consider this compound relevant in our study). Similar situations occurred for other compounds tentatively identified, as metformin, pirantel, and the sweetener saccharin, due to the lack of reference standard (circled plus in Table 1).

3. Searching for metabolites

A detailed discussion is made on the particular case of losartan, an angiotensin II receptor antagonist drug used mainly to treat high blood pressure (hypertension), and its metabolites identified making use of the common fragmentation pathway.

Losartan was confirmed to be present in all the samples analyzed. The LE spectrum in ESI+ showed an abundant chromatographic peak at the expected retention time (10.13 min) for this pharmaceutical (m/z 423.1695, -1.2 ppm mass error) and also presented the typical isotopic pattern of a chlorine atom ($C_{22}H_{24}N_6OCl$, expressed as protonated molecule). The expected fragment ion at m/z 405.1590 ($C_{22}H_{22}N_6Cl$, -1.0 ppm), due to a loss of water, was also observed in the LE spectrum, as well as a minor fragment ion at m/z 377.1528 ($C_{22}H_{22}N_4Cl$, -1.3 ppm), corresponding to a N_2 loss from m/z 405. The HE spectrum showed a predominant ion at m/z 207.0916 ($C_{14}H_{11}N_2$) and minor peaks at m/z 235.0979 ($C_{14}H_{11}N_4$), 192.0813 ($C_{14}H_{10}N$), 190.0661 ($C_{14}H_8N$), and 180.0805 ($C_{13}H_{10}N$), all with mass errors below 4 ppm. These fragments correspond to the fragmentation at the benzylic carbon (CH_2 near to the two rings) (m/z 235) with losses of N_2 (m/z 207), N_3H (m/z 192), or N_3H_3 (m/z 190). Losartan was also detected in negative mode at m/z 421.1552, showing fragment ions at m/z 187.0631 ($C_8H_{12}N_2OCl$), 179.0868 ($C_{14}H_{11}$), 157.0529 ($C_7H_{10}N_2Cl$), and 127.0066 ($C_5H_4N_2Cl$). With all this information, the compound detected in the samples was unequivocally confirmed to be losartan.

The presence of three additional chromatographic peaks at 8.25, 9.29, and 10.47 min in all the three XICs at m/z 207, 235, and 190 performed at the HE function in ESI+ (for losartan, retention time 10.13 min) suggested that the three compounds were chemically related with this pharmaceutical (Fig. 4). The common fragmentation pathway strategy has been successfully applied by our group in the elucidation of metabolites of the new psychoactive substance methylenedioxypropylvalerone (MDPV) and of degradation products of cocaine and benzoylecgonine [27, 28]. This encouraged us to investigate the identity of the potential metabolites/transformation products of losartan from the data provided by QTOF MS.

Firstly, the LE spectra (all showing the isotopic distribution corresponding to a chlorine atom) of these three possible metabolites were studied in more detail. In relation to the peak at 10.47 min, the protonated molecule corresponded to m/z 437.1490 ($C_{22}H_{22}N_6O_2Cl$, -0.7 ppm), which might be attributed to an oxidation of losartan. This compound might correspond to losartan carboxylic acid, the main metabolite reported in the literature [29].

Regarding the peaks at 8.25 and 9.29 min, accurate masses of m/z 439.1643 and 439.1661 were obtained, respectively, both corresponding to the same empirical formulae $C_{22}H_{24}N_6O_2Cl$ with mass errors lower than 3 ppm. This implies the presence of an extra oxygen atom with respect to losartan. Hence, these two compounds might correspond to hydroxylated metabolites of losartan.

The following step consisted on the elucidation of the structure of these three potential metabolites. For this purpose, MS/MS experiments were performed at different collision energies (10–50 eV).

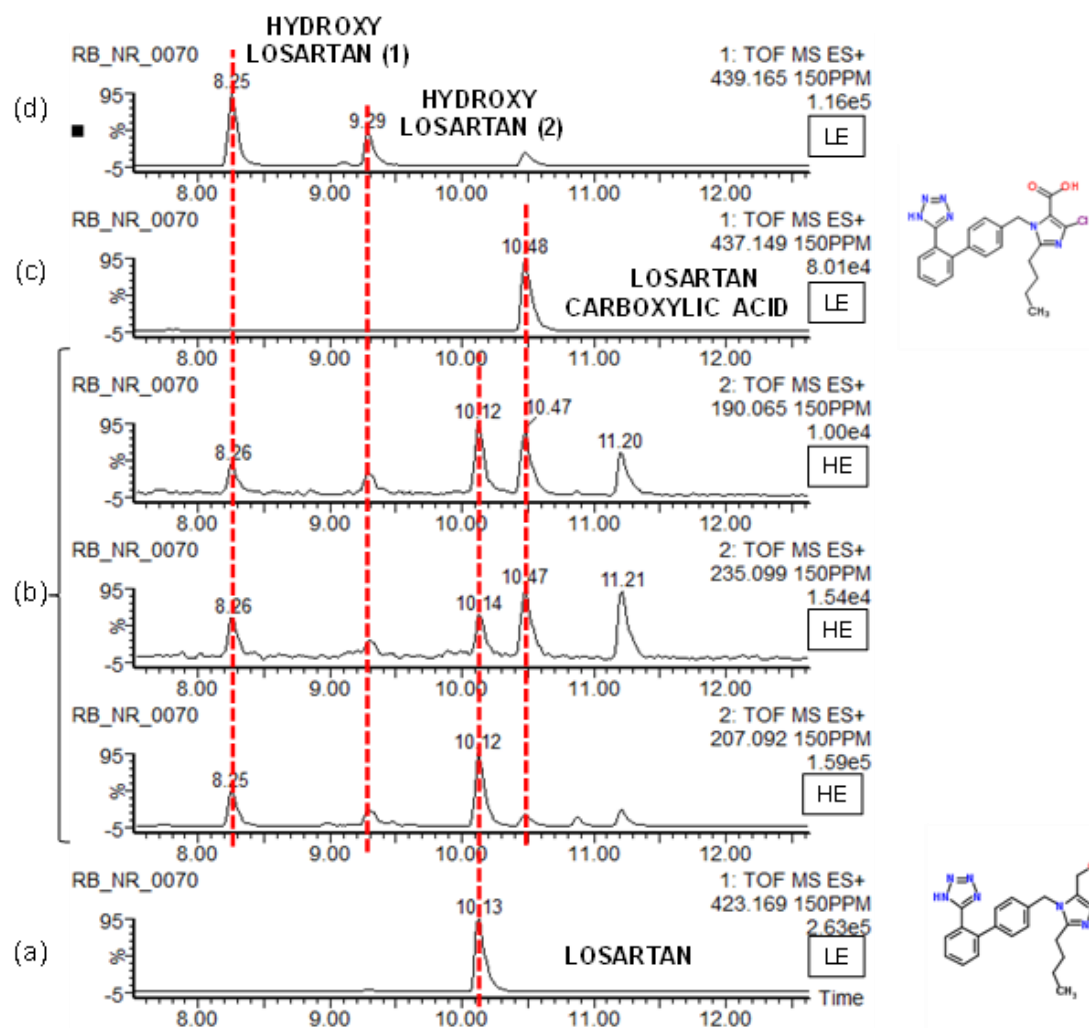


Fig. 4 Common fragmentation pathway strategy applied for the detection of metabolites of losartan. (a) eXtracted Ion Chromatograms (XICs) at 150 ppm mass window for $[M+H]^+$ of losartan in LE function and (b) main fragments (m/z 207, 235, 190) in HE function. (c) XICs at 150 ppm mass window for $[M+H]^+$ of possible carboxylic acid metabolite in LE function. (d) XICs at 150 ppm mass window for $[M+H]^+$ of possible hydroxylated metabolites of losartan in LE function

Regarding the peak at 10.47 min (m/z 437.1490, $C_{22}H_{22}N_6O_2Cl$, -0.7 ppm), five common product ions with losartan were observed at m/z 235.0982, 207.0919, 192.0816, 190.0654, and 180.0809. This compound was also detected in negative mode, with m/z 435.1327 ($C_{22}H_{22}N_6O_2Cl$, -2.1 ppm). At the lowest collision energy, a product ion was observed at m/z 391.1429 ($C_{21}H_{20}N_6Cl$, -2.3 ppm), corresponding to a loss of CO_2 . When the collision energy was increased, product ions at m/z 363.1375 ($C_{21}H_{20}N_4Cl$, this is a N_2 loss

from 391, -0.3 ppm), 157.0530 ($C_7H_{10}N_2Cl$, -1.9 ppm), and 113.9989 ($C_4H_3N_2Cl$, 3.5 ppm) were observed. All these fragments fitted with the chemical structure of losartan carboxylic acid. After this careful evaluation and well-supported tentative identification, the reference standard of losartan carboxylic acid was acquired and injected, allowing the ultimate confirmation of this metabolite in the samples.

Regarding the hydroxylated metabolite 1 (8.25 min, m/z 439.1643, $C_{22}H_{24}N_6O_2Cl$, -1.4 ppm), two minor product ions were observed at 10 eV at m/z 421.1536 ($C_{22}H_{22}N_6OCl$) and 385.1772 ($C_{22}H_{21}N_6O$), resulting from a loss of water and subsequent loss of hydrochloric acid. At higher collision energies, product ions at m/z 235.0978 ($C_{14}H_{11}N_4$), 207.0915 ($C_{14}H_{11}N_2$), 192.0806 ($C_{14}H_{10}N$), 190.0649 ($C_{14}H_8N$), and 180.0803 ($C_{13}H_{10}N$) were obtained (Fig. 5a). Thus, ESI+ fragmentation advises about what is happening near to the tetrazol group but not about the location of the hydroxyl group. Accordingly, these fragments seem to indicate that the hydroxylation has occurred in the imidazole part. In negative mode, product ions at m/z 203.0585 ($C_8H_{12}N_2O_2Cl$) and 173.0488 ($C_7H_{10}N_2OCl$) were observed at 20 eV. When the collision energy was increased up to 50 eV, product ions at m/z 155.0384 ($C_7H_8N_2Cl$), 127.0060 ($C_5H_4N_2Cl$), 113.9991 ($C_4H_3N_2Cl$), and 100.9912 ($C_3H_2N_2Cl$) were observed (Fig. 5b). Considering all this information, the $-OH$ group could be placed in the benzimidazole part, being an N -oxide the most plausible candidate. Otherwise, if the hydroxyl group was located in the alkylic chain, a second loss of water should in principle be observed in ESI+.

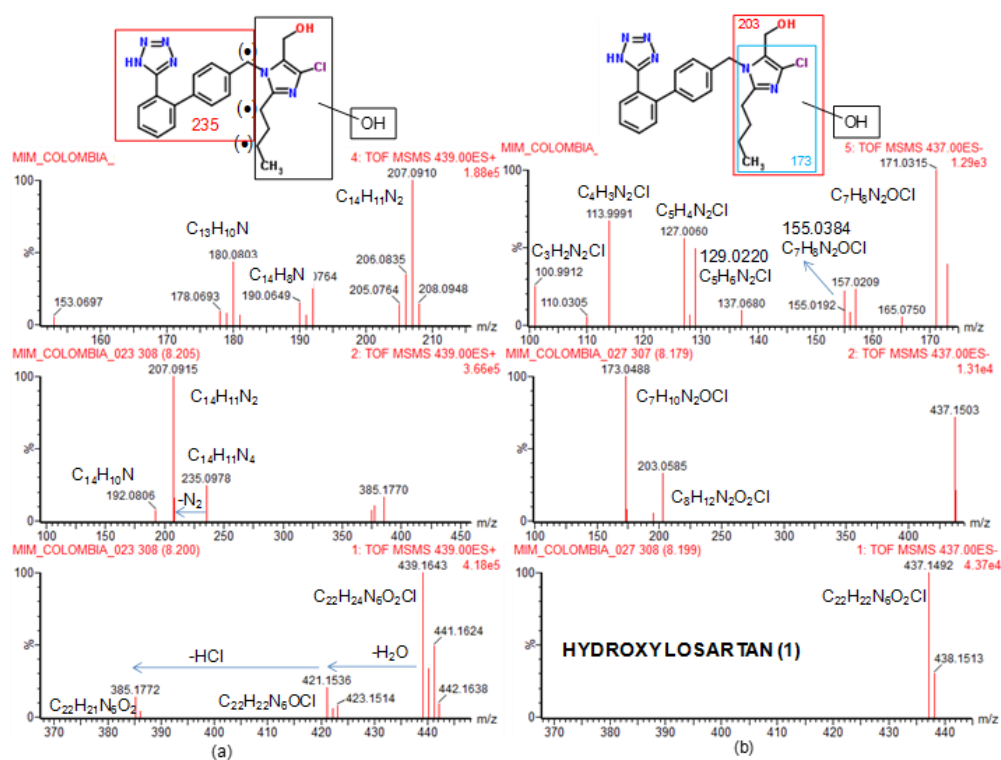


Fig. 5 MS/MS spectra at 10 eV (bottom), 20 eV (middle), and 50 eV (top) for hydroxylated metabolite 1 (8.25 min) in (a) ESI+ and (b) ESI- (dot) indicates the possible hydroxylation sites, as reported in the literature

Regarding the hydroxylated metabolite 2 (9.29 min, m/z 439.1645, $C_{22}H_{24}N_6O_2Cl$, -0.9 ppm), two important fragment ions were observed at 10 eV (ESI+) at m/z 421.1538 ($C_{22}H_{22}N_6OCl$) and 403.1435 ($C_{22}H_{20}N_6Cl$) corresponding to two consecutive losses of water. At higher collision energies, the common fragment ions at m/z 235.0983, 207.0920, 192.0811, 190.0659, and 180.0808 were observed. The spectra in negative mode were different from those of losartan and the other hydroxylated metabolite (e.g., respectively, m/z 187/157 or 203/173 were not seen), and only two product ions at m/z 100.9908 ($C_3H_2N_2Cl$) and 131.0008 ($C_4H_4N_2OCl$) were observed. The presence of the $-OH$ group on the other N of the imidazole group would not explain the easy loss of water observed in ESI+. Another possibility would be the hydroxyl group to be located in the benzylic carbon, although this would hamper (but not prevent) the formation of the ions 235/207 in ESI+.

After a literature search [29], three hydroxylated candidates were found (see the hydroxylated reported sites, marked as dot in Fig. 5): two metabolites are hydroxylated in the alkylic chain and the other in the benzylic carbon. However, on the basis of the MS/MS fragmentation, it was not possible to unequivocally locate the position of the $-OH$ group. It would be very interesting to perform NMR to help identify the structure of these compounds. However, this was not possible due to the expected low concentrations of the discovered TPs together with the need for on-line coupling of UHPLC separations to NMR spectroscopy. Unfortunately, this type of instrumentation is not easily available in our area.

It is interesting to remark that losartan, its carboxylic acid, and the two hydroxylated metabolites were found in all water samples analyzed (ten surface, seven effluent WW, seven influent WW).

Conclusions

The potential of UHPLC-QTOF MS for large screening of more than 1,000 licit and illicit drugs, even without the need of having all reference standards available, has allowed the detection of many of these compounds in water samples from the area of Bogotá. The screening performed in urban wastewater and surface water confirmed the presence of emerging contaminants, mainly pharmaceuticals, in the samples. This work provides new information on the occurrence of these compounds and metabolites in the Colombian water cycle. The availability of full-spectrum acquisition accurate mass QTOF data will also facilitate retrospective analysis of other contaminants in the future, not considered in this initial screening, if required. The results obtained in this first study may help Colombian institutions to select priority contaminants for future actions. Thus, target methodologies, typically based on LC-MS/MS, will need to be set up in the near future for accurate and sensitive quantification of the contaminants selected on the basis on the information provided in the present paper.

Acknowledgments

The authors are very grateful to Nubia Irley León and colleagues of Acueducto de Bogotá (Bogotá, Colombia) for providing wastewater samples, Natalia Gonzalez for preparing the wastewater samples, Jose Roberto Vergas for collecting and processing the surface water samples, and Carlos Sales Martinez for making Fig. 1. Furthermore, the collaboration of Juan V. Sancho and Oscar J. Pozo of the Hospital del Mar Medical Research Institute, Bioanalysis Research Group (Barcelona, Spain), for the interpretation of the fragmentation pathways of the losartan metabolites is highly appreciated.

This work has been developed under financial support provided by the Plan Nacional de I+D+I, Ministerio de Economía y Competitividad (Project ref. CTQ2012-36189) and the University Antonio Nariño, Colombia (Project ref. 20141082). The authors also acknowledge the financial support from Generalitat Valenciana (Group of Excellence Prometeo 2009/054, Prometeo II 2014/023; Collaborative Research on Environment and Food Safety ISIC/2012/016).

R. Bade acknowledges the European Union for his Early Stage Researcher (ESR) contract as part of the EU-International Training Network SEWPROF (Marie Curie- PEOPLE Grant #317205).

References

1. Pal A, He Y, Jekel M, Reinhard M, Yew-Hoon Gin K (2014) Emerging contaminants of public health significance as water quality indicator compounds in the urban water cycle. *Environm Int* 71:46-62
2. Richardson SD, Ternes TA (2014) Water Analysis: Emerging Contaminants and Current Issues. *Anal Chem* 86:2813-2848
3. Fatta-Kassinos D, Meric S, Nikolaou A (2011) Pharmaceutical residues in environmental waters and wastewater: current state of knowledge and future research. *Anal Bioanal Chem* 399:251-275
4. Gros M, Petrovic M, Ginebreda A, Barceló D (2010) Removal of pharmaceuticals during wastewater treatment and environmental risk assessment using hazard indexes. *Environm Int* 36:15-26
5. Thomas KV, Araújo da Silva FM, Langford KH, Leão de Souza A, Nizzeto L, Waichman AV (2014) Screening for selected human pharmaceuticals and cocaine in the urban streams of Manaus, Amazonas, Brazil. *JAWRA* 50:302-308
6. De Paula VS, Diniz-Mendes L, Villar LM, Luz SLB, Silva LA, Jesus MS, da Silva NMVS, Gaspar AMC (2007) Hepatitis A virus in environmental water samples from the Amazon Basin. *Water Res* 41:1169-1176
7. Montagner CC, Jardim WF (2011) Spatial and seasonal variations of pharmaceuticals and endocrine disruptors in the Atibaia River, São Paulo State (Brazil). *J Braz Chem Soc* 22:1452-1462
8. Gracia-Lor E, Martínez M, Sancho JV, Peñuela G, Hernández F (2012) Multi-class determination of personal care products and pharmaceuticals in environmental and wastewater samples by ultra-high performance liquid-chromatography-tandem mass spectrometry. *Talanta* 99:1011-1023
9. Boxall ABA, Rudd MA, Brooks BW, Caldwell DJ, Choi K, Hickmann S, Innes E, Ostapyk K, Staveley JP, Verslycke T, Ankley GT, Beazley KF, Belanger SE, Berninger JP, Carriquiriborde P, Coors A, DeLeo PC, Dyer SD, Ericson JF, Gagné F, Giesy JP, Gouin T, Hallstrom L, Karlsson MV, Larsson DGJ, Lazorchak JM, Mastrocco F, McLaughlin A, McMaster ME, Meyerhoff RD, Moore R, Parrott JL, Snape JR, Murray-Smith R, Servos MR, Sibley PK, Straub JO, Szabo ND, Topp E, Tetreault GR, Trudeau VL, van der Kraak G (2012) Pharmaceuticals and Personal Care Products in the Environment: What Are the Big Questions? *Environ Health Perspect* 120:1221-1229
10. Schymanski EL, Singer HP, Longrée P, Loos M, Ruff M, Stravs MA, Ripollés Vidal C, Hollender J (2014) Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. *Environ Sci Technol* 48:1811-1818
11. Hernández F, Ibáñez M, Bade R, Bijlsma L, Sancho JV (2014) Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry. *Trends Anal Chem* 63:140-157

12. Díaz R, Ibáñez M, Sancho JV, Hernández F (2012) Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF-MS. *Anal Meth* 4:196-209
13. Hernández F, Bijlsma L, Sancho JV, Díaz R, Ibáñez M (2011) Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography–quadrupole-time-of-flight-mass spectrometry. *Anal Chim Acta* 684: 96-106
14. Laks S, Pelander A, Vuori E, Ali-Toppa E, Sippola E, Ojanpera I (2004) Analysis of street drugs in seized material without primary reference standards. *Anal Chem* 76: 7375-7379
15. Hernández F, Portolés T, Ibáñez M, Bustos-López MC, Díaz R, Botero-Coy AM, Fuentes CL, Peñuela G (2012) Use of time-of-flight mass spectrometry for large screening of organic pollutants in surface waters and soils from a rice production area in Colombia. *Sci Total Environ* 439: 249-259
16. MassBank Norman. High Resolution Mass Spectral Database. <http://massbank.ufz.de/MassBank/> Accessed 29 Jan 2015
17. Metlin. Metabolite and tandem MS database. <http://metlin.scripps.edu/index.php> Accessed 29 Jan 2015
18. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J (2014) Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ Sci Technol* 48:2097-2098
19. Hernández F, Ibáñez M, Gracia-Lor E, Sancho JV (2011) Retrospective LC–QTOF-MS analysis searching for pharmaceutical metabolites in urban wastewater. *J Sep Sci* 34:3517–3526
20. López JJ, Dennis R, Moscoso SM (2009) Estudio sobre la Automedicación en una localidad de Bogotá. *Revista de Salud Pública* 11:432-442
21. Rodríguez J, Ruiz F, Peñaloza E, Eslava J, Gómez LC, Sanchez H, Amaya JL, Arenas R, Botiva Y. (2009) Encuesta Nacional de Salud 2007. ISBN 978-958-716-220-2
22. Machado-Alba JE, González-Santos DM (2009) Dispensación de antibióticos de uso ambulatorio en una población colombiana. *Revista de Salud Pública* 11:734-744
23. United Nations Office on Drugs and Crime. https://www.unodc.org/documents/colombia/2014/Julio/Estudio_de_Consumo_UNODC.pdf Accessed 26 Jan 2015

24. Bijlsma L, Beltrán E, Boix C, Sancho JV, Hernández F (2014) Improvements in analytical methodology for the determination of frequently consumed illicit drugs in urban wastewater. *Anal Bioanal Chem* 406:4261–4272
25. Ort C, van Nuijs ALN, Berset J-D, Bijlsma L, Castiglioni S, Covaci A, de Voogt P, Emke E, Fatta-Kassinos D, Griffiths P, Hernández F, González-Mariño I, Grabic R, Kasprzyk-Hordern B, Mastroianni N, Meierjohann A, Nefau T, Östman M, Pico Y, Racamonde I, Reid M, Slobodnik J, Terzic S, Thomaidis N, Thomas KV (2014) Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis. *Addiction*, 109;1338-1352
26. Pedrouzo M, Borrull F, Pocurull E, Marcé RM (2011) Drugs of abuse and their metabolites in waste and surface waters by Liquid Chromatography-tandem Mass Spectrometry. *J Sep Sci* 34:1091–1101
27. Ibáñez M, Pozo OJ, Sancho JV, Orengo T, Haro G, Hernández F (2015) Discovering human urinary metabolites of methylenedioxypropylamphetamine (MDPV) by Liquid Chromatography/Quadrupole Time-of-Flight Mass Spectrometry. *Forensic Toxicol* submitted
28. Bijlsma L, Boix C, Niesen WMA, Ibáñez M, Sancho JV, Hernández F (2013) Investigation of degradation products of cocaine and benzoylecgonine in the aquatic environment. *Sci Total Environ* 443:200–208
29. PGKB Pharmacogenomics. <https://www.pharmgkb.org/pathway/PA164713428#> Accessed 29 Jan 2015

2.2.2 Scientific Article 3

**SCREENING OF PHARMACEUTICALS AND ILLICIT
DRUGS IN WASTEWATER AND SURFACE WATERS OF
SPAIN AND ITALY BY HIGH RESOLUTION MASS
SPECTROMETRY USING UHPLC-QTOF MS AND
LC-LTQ-ORBITRAP MS**

Richard Bade, Nikolaos I. Rousis, Lubertus Bijlsma, Emma Gracia-Lor, Sara Castiglioni, Juan V. Sancho, Felix Hernandez

Analytical and Bioanalytical Chemistry (2015) 407:8979-8988



RESEARCH PAPER

Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS

Richard Bade¹ · Nikolaos I. Rousis² · Lubertus Bijlsma¹ · Emma Gracia-Lor² · Sara Castiglioni² · Juan V. Sancho¹ · Felix Hernandez¹

Received: 22 June 2015 / Revised: 10 September 2015 / Accepted: 17 September 2015 / Published online: 28 September 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract

The existence of pharmaceuticals and illicit drugs (PIDs) in environmental waters has led many analytical chemists to develop screening methods for monitoring purposes. Water samples can contain a huge number of possible contaminants, commonly at low concentrations, which makes their detection and identification problematic. Liquid chromatography coupled with high resolution mass spectrometry (LC-HRMS) has proven itself effective in the screening of environmental contaminants. The present work investigates the use of the most popular HRMS instruments, quadrupole time-of-flight and linear trap quadrupole-Orbitrap, from two different laboratories. A suspect screening for PIDs was carried out on wastewater (influent and effluent) and surface water samples from Castellón, Eastern Spain, and Cremona, Northern Italy, incorporating a database of 107 PIDs (including 220 fragment ions). A comparison between the findings of both instruments and of the samples was made which highlights the advantages and drawbacks of the strategies applied in each case. In total, 28 compounds were detected and/or identified by either/both instruments with irbesartan, valsartan, benzoylecgonine and caffeine being the most commonly found compounds across all samples.

Keywords: Liquid chromatography, Screening, Pharmaceuticals, Illicit drugs, High resolution mass spectrometry

Introduction

The presence of pharmaceuticals and illicit drugs (PIDs) in environmental waters has become more apparent in the past decade due to the improvements in selectivity and sensitivity of modern analytical techniques. PIDs are continuously excreted or disposed into the sewer systems as the unaltered parent compound or as metabolites [1]. Compounds of the most concern in the environment are those found with high concentrations and those that have been proven detrimental to the aquatic life; in fact, for several antibiotics, hormones and pharmaceuticals, some effects have been shown [2–4]. Various pharmaceuticals, antibiotics and illicit drugs and metabolites such as acetaminophen, carbamazepine, diclofenac, ciprofloxacin and benzoylecgonine are frequently detected in waters and are considered a potential threat [2, 5, 6].

Nowadays, data processing is often the most time-consuming step when screening a large number compounds. A fast screening method, referring to the easy searching of PIDs by software following full scan acquisition by high resolution mass spectrometry (HRMS), is therefore of great interest in order to gain a complete overview of possible contaminants and their fate in the aquatic environment. Another time-consuming step is sample preparation. Important improvements in sensitivity and specificity of recent instrumentation have facilitated the direct injection of the samples [7, 8].

HRMS instruments, such as quadrupole time-of-flight (QTOF) MS and Orbitrap MS, provide high quality data by combining full spectrum mass data with high mass resolution and mass accuracy [9]. Powerful deconvolution software is required to facilitate sample analysis for these instruments. Each manufacturer has its own software, such as ChromaLynx (Waters) and TraceFinder (ThermoScientific). These systems allow post-target analysis, whereby the presence of an unlimited number of contaminants can be investigated through the addition of empirical or theoretical compound databases, without depending on the pre-selection of analytes or having reference standards available [10]. In the processing of numerous samples and compounds, these programs undeniably save time and hence are a necessity for a large screening of environmental contaminants.

It is difficult to compare the individual HRMS mass analysers (i.e. TOF and Orbitrap), but each hold distinct advantages. The main advantage of Orbitrap is its high mass resolving power ($>100,000$ full width at half maximum (FWHM)). In complex environmental matrices, co-elution of analytes with matrix interferences can result in ambiguous peaks. By utilising the ultra-high resolution capabilities, compounds with a very similar exact mass can be easily differentiated [11]. However, its main drawback is its slower scanning speed and inverse relationship to resolution. Conversely, the higher scanning speed of TOF enables it to be used with the highly efficient separation technique ultra-high performance liquid chromatography (UHPLC), resulting in shorter run times and improved sensitivity [10].

To achieve better performance such as higher sensitivity and mass resolving power, manufacturers have implemented hybrid mass spectrometers, two of which are used in the current work: QTOF and linear trap quadrupole-Orbitrap (LTQ-Orbitrap). The UHPLC capabilities are exemplified when QTOF is employed in

MSE mode, which involves the simultaneous acquisition of exact mass data at low and high collision energy [12]. At low energy, information pertaining to the non-fragmented (de)protonated molecule is obtained, whilst at high energy, information regarding the fragments is attained. This means that all required information is obtained in just one injection. LTQ-Orbitrap combines the tandem mass spectrometric capability of the LTQ with the high resolution and mass accuracy capability of the Fourier transform (FT) Orbitrap [13]. LTQ-Orbitrap is operated in data-dependent acquisition (DDA) mode whereby MS and MS_n spectra are obtained simultaneously; however, the number of compounds able to be acquired is limited and therefore may require subsequent injections. Some studies have been made on the identification of organic pollutants in water by QTOF MS [12, 14–16] and LTQ-Orbitrap MS [17–19].

This work derives from a collaborative study between two laboratories within the ITN Marie Curie SEWPROF network [20]. The aim was to perform a screening of PIDs in waste- and environmental waters using two different liquid chromatography coupled with high resolution mass spectrometry (LC-HRMS) systems (UHPLC-QTOF MS and Capillary-LC-LTQ-Orbitrap MS) available in our laboratories. Normally, a given laboratory would use just one of the instruments, but as both have a great capability for screening and identification, in this work, we pursued the comparison of data from the same analysed samples (samples collected from Spain and Italy). In addition, the complementary use allows true confirmation of the compound in the sample [9]. Although the same compounds were initially searched, the different acquisition modes may additionally detect other contaminants and widen the scope of the screening. It is worth noting that a comparison of the instruments or methods used was not pursued, rather their use to ratify the findings of the other instrument. In this way, on the basis of the results of the present paper, the methodologies applied by each laboratory could be validated by the results obtained by the other laboratory in terms of qualitative analysis. When performing the screening, four criteria were used based on the level of confidence and on the availability of reference standards. Numerous compounds were detected and identified from samples of different origin and matrix composition (influent and effluent urban wastewater, surface water). All samples, independently of they were collected from Spain or from Italy, were analysed by both HRMS instruments (Orbitrap in Milan and TOF in Castellón), which supports the feasibility and robustness of high resolution screening of PIDs in complex environmental samples.

Experimental

Chemicals and reagents

For information regarding the reference standards used for QTOF and LTQ-Orbitrap analysis (see Electronic Supplementary Material (ESM), Section 2.1).

HPLC-grade methanol (MeOH), ammonia solution (25 %) and formic acid (HCOOH, 98–100 %) were acquired from Scharlau (Barcelona, Spain) and acetonitrile for LC-MS from Riedel de Haen (Seelze, Germany). HPLC-grade water was obtained by purifying demineralised water in a Milli-Q plus system from Millipore (Bedford, MA, USA). Solid phase extraction (SPE) cartridges used were Oasis HLB 3 mL (60 mg) from Waters (Milford, MA, USA). Mixed cellulose ester membrane filters (0.45 μm) and 1.6 μm GF/A glass microfiber filters were purchased from Whatman (Kent, UK).

Samples

Influent and effluent wastewater

In total, 18 wastewater samples (nine influent wastewater (IWW) and effluent wastewater (EWW)) were taken from a wastewater treatment plant (WWTP) in Castellón, Eastern Spain and Cremona, Northern Italy. The 24-h composite IWW and EWW samples were collected over seven consecutive days in March 2014 (Castellón) and two consecutive days in May (Cremona) in high-density polystyrene bottles and directly transported to the lab where they were immediately filtered and stored in the dark at $-20\text{ }^{\circ}\text{C}$ until analysis.

Surface waters

Five grab samples were taken from the Castellón region, Eastern Spain: Almenara, Burriana Clot, Nules and two sites in Albufera Natural Park. All samples were stored in high-density polystyrene bottles at $4\text{ }^{\circ}\text{C}$ for less than 48 h, until sample treatment.

Sample treatment

Prior to SPE, Spanish WW samples were vacuum filtered through 0.45 μm mixed cellulose ester membrane filters, and the Italian samples were filtered through 1.6 μm GF/A glass microfiber filters. Surface water samples were not filtered. Subsequently, 100 mL of the samples (IWW was four times diluted with Milli-Q water, i.e. 25 mL sample in 100 mL) was extracted using SPE cartridges (Oasis HLB), which were first conditioned by washing and rinsing with 6 mL of MeOH and 6 mL of Milli-Q water. Samples were percolated through the cartridges by gravity (flow rate around 3 mL min^{-1}) and vacuum dried for approximately 15 min. After elution with methanol, extracts were evaporated to dryness at $35\text{ }^{\circ}\text{C}$ under a gentle stream of nitrogen and reconstructed in 1 mL of 10 % methanol aqueous solution [21]. Two 0.5-mL aliquots were taken from the extract, one each for the QTOF and Orbitrap analyses. UHPLC-MS QTOF analyses were performed by

injecting 20 μL of the final extract into the system. Capillary-LC-LTQ-Orbitrap analyses were performed by injecting 2 μL of the final extract into the system.

Instrumentation

Ultra high-performance liquid chromatography-QTOF-mass spectrometry

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (XEVO G2 QTOF, Waters Micromass, Manchester, UK), using a Z-Spray ESI interface operating in positive ion mode.

The chromatographic separation was performed using an Acquity UPLC BEH C18 1.7 μm particle size column 100 \times 2.1 mm (Waters) at a flow rate of 300 $\mu\text{L}/\text{min}$. The mobile phases used were A = H₂O and B = MeOH (both with 0.01 % HCOOH). The total run time was 18 min. The initial percentage of B was 10 %, which was linearly increased to 90 % in 14 min, followed by a 2-min isocratic period and, then, returned to initial conditions during 2 min. Nitrogen was used as drying gas and nebulizing gas.

MS data were acquired over an m/z range of 50–1,000. A capillary voltage of 0.7 kV and cone voltage of 20 V were used. Collision gas was argon 99.995 % (Praxair, Valencia, Spain). The desolvation temperature was set to 600 °C and the source temperature to 135 °C. The column temperature was set to 40 °C.

For MS E experiments, two acquisition functions with different collision energies were created: the low energy function (LE), selecting a collision energy of 4 eV, and the high energy (HE) function, with a collision energy ramp ranging from 15 to 40 eV in order to obtain a greater range of fragment ions. The total scan time was 0.4 s.

QTOF MS data was processed using ChromaLynx XS application manager (within MassLynx v 4.1; Waters Corporation). The following parameters were used: target RT tolerance (1.50 min) from empirical retention time (obtained from injecting standards), mass tolerance 0.020 Da (for positive ID 0.010 Da), peak width at 5 % height 6 s, peak-to-peak baseline noise 1,000 and threshold absolute area: 250 for SW and 500 for IWW and EWW.

Liquid chromatography-LTQ-Orbitrap mass spectrometry

Mass spectrometry analysis was performed on a high resolution LTQ-Orbitrap XL instrument, interfaced to the HPLC system through a desorption electrospray ionisation (DESI) source and used in nano-spray mode (Prosolia, Indianapolis, IN, USA). The instrument was operated at 60,000 resolution for full MS scans and at 15,000 resolution for MS₂ scans, after CID fragmentation in the linear ion trap cell. The scan time was 0.85 s and the cycle time was 4.8 s. Every sample was analysed in a DDA mode, in which the first data event was a full scan MS, whilst the following four data events were MS₂ scans of the four most intense ions recorded in

the first event. A subsequent injection in MS/MS mode was made to find further fragments, based on target lists (see Section 3.1).

The chromatographic separation was performed using a Zorbax SB-C18 5 μm particle size column 150 \times 0.5 mm (Agilent Technologies) at a flow rate of 10 $\mu\text{L}/\text{min}$. The mobile phases used were A: H₂O (0.1 % HCOOH) and B: acetonitrile. The total run time was 46 min. The initial percentage of B was 1 % and was linearly increased to 99 % in 24 min, followed by a 10-min isocratic period and then reduced back to initial conditions in 2 min which were kept for 10 min (post-run) to equilibrate the column.

Processing of the Orbitrap data was made using TraceFinder EFS and QualBrowser software (Thermo Scientific). The following parameters were used: [M+H]⁺ intensity (5,000), signal-to-noise ratio threshold (=5), retention time window override for confirmation (90 s), fragment intensity (1,000 with 10 ppm mass tolerance); isotopic pattern (60 % fit threshold, 5 ppm mass deviation, 10 % intensity deviation); and library search (20 % score threshold) using the Toxicology HCD 30-70-110 eV Library of Thermo Scientific. This library contained the information of 1,700 compounds, but not all PID compounds selected in our list were included in the library.

Criteria for detection/identification

Up to four levels were considered for detection and identification of the compounds based on the availability of reference standard and on the reliability of the data obtained. These levels were derived from a recent article exploring the means of identification of small molecules [22] and the European Commission Decision 2002/657/EC for the confirmation of contaminants in samples of animal origin, commonly employed in environmental studies. The criteria used included the presence of the protonated molecule at accurate mass, fragment ion(s) at accurate mass, retention time and isotopic pattern.

Selected criteria for QTOF MS

The standards injected on each instrument were those available at the respective laboratories (see ESM Section 2.1). They may be different; however, as explained above, a comparison between the labs is not proposed.

Standard available (11 illicit drug and 61 pharmaceutical standards, see ESM Section 2.1)

1. Detection: Protonated molecule [M+H]⁺ at accurate mass (<5 ppm) and retention time in agreement with reference standard (± 2.5 %)
2. Identification: Protonated molecule [M+H]⁺ at accurate mass (<5 ppm), at least one fragment ion, commonly in HE (<5 ppm), retention time agreement with reference standard (± 2.5 %)

Standard not available

3. Tentative detection (potential positive): Presence in the extracted ion chromatogram of the protonated molecule $[M+H]^+$ at accurate mass (<10 ppm), typically in the LE function
4. Tentative identification: Protonated molecule $[M+H]^+$ at accurate mass (<10 ppm) and at least one fragment ion, commonly in HE function (<10 ppm), in agreement with data reported from literature or online databases

Selected criteria for LTQ-Orbitrap MS

Standard available (25 illicit drug and 27 pharmaceutical standards, see ESM Section 2.1)

1. Detection: Protonated molecule $[M+H]^+$ at accurate mass (<5 ppm), retention time agreement with standard ($\pm 2.5\%$), isotopic pattern (60 % fit threshold, 5 ppm mass deviation, 10 % intensity deviation)
2. Identification: Protonated molecule $[M+H]^+$ at accurate mass (<5 ppm, intensity $>5,000$), at least one fragment ion in MS/MS (intensity $>1,000$, <10 ppm), retention time agreement with standard ($\pm 2.5\%$), isotopic pattern (60 % fit threshold, 5 ppm mass deviation, 10 % intensity deviation); library search (20 % score threshold) using the Toxicology HCD 30-70-110 eV Library.

Standard not available

3. Tentative detection (potential positive): Protonated molecule $[M+H]^+$ at accurate mass (<10 ppm, intensity $>5,000$)
4. Tentative identification: Protonated molecule $[M+H]^+$ at accurate mass (<10 ppm intensity $>5,000$), at least one fragment ion (intensity $>1,000$, <10 ppm mass tolerance), comparable with literature or online databases, isotopic pattern (60 % fit threshold, 5 ppm mass deviation, 10 % intensity deviation)

Results and discussion

The methodologies followed for these analyses are based on the generic configuration employed by each laboratory. We acknowledge that the Orbitrap configuration is not optimal for screening of PIDs in water, as its use is primarily in the field of proteomics, which justifies the use of a capillary column and DESI source. Therefore, we do not intend to show a comprehensive comparison of the instruments but to determine whether both laboratories and instruments can detect and identify the same compounds using their individual workflows.

Furthermore, in this work, we do not set out to perform a wide-scope screening of a huge number of compounds but rather to focus on a limited number of well-known, widely detected PIDs in wastewater to test the usefulness of the different approaches and instruments used and also to compare the countries.

Selection of PIDs

The compounds to be investigated were initially based on a database similar to that made by Diaz et al. [23] but focussing solely on pharmaceuticals and illicit drugs. The database contained the compound name, associated fragments (if known), retention time (if standard was available) and molecular formula. A qualitative validation was previously performed using a similar analytical procedure based on LC-QTOF MS measurements, and the empirical screening detection limits (SDL) and limits of identification (LOI) in different type of water samples were established for some of the compounds included in the present work. For most compounds evaluated, SDLs were 0.02 or 0.1 µg/L [24, 25]. For compounds where no standards were available, the database included data regarding fragment ions and molecular formula from literature or online databases such as MassBank [26].

To ensure national significance, compounds found from previous studies of these regions were investigated and added to the database (Spain [27, 28]; Italy [29–31]). In total, 107 PIDs and 220 fragment ions were incorporated in the database (see ESM Table S1).

As stated in Section 2.4.2, each sample was injected twice in the Orbitrap. The first injection enabled full scan acquisition. In this data, the original database of 107 PIDs was searched using TraceFinder and potential positive compounds were then selected, based on retention time and exact mass, as unfortunately the expected fragment ion information of the targeted compounds was not amongst the four most intense ions detected within the data events. These “positives” were then grouped into target lists, according to whether they were found in the Spanish (40 compounds) or Italian (52 compounds) samples (see ESM Tables S2 and S3). The second injection incorporated these target lists, thereby enabling their product ions to be seen.

Within the software of the QTOF and Orbitrap, a threshold was given for intensity and mass error, outside which peaks were disregarded as noise. The majority of compounds were able to be automatically identified or detected thanks to the ChromaLynx and TraceFinder software. However, in some cases, some compounds were only able to be detected (i.e. no fragment ions found). This demonstrated the complementariness of the instruments: when one instrument was solely able to detect a compound (only the precursor ion found), the other was able to at least tentatively identify the compound with fragments. The raw data of both instruments was manually inspected in these situations to ratify the fragments (see Table 1). This secondary examination also reduced the possibility that one of the software had erroneously eliminated some compounds. The raw data investigations were only required for a small number of compounds, which is testament to the utility of the software. However, by decreasing the identification thresholds used, the need for manual inspection could be avoided, which is especially recommended for wide-scope screening methods incorporating thousands of compounds.

Table 1: All compounds detected/identified in wastewater and surface water samples of Spain and Italy

Compound	Spain IWW (n=7)		Spain EWW (n=7)		Spain SW (n=5)		Italy IWW (n=2)		Italy EWW (n=2)	
	QTOF	Orbitrap	QTOF	Orbitrap	QTOF	Orbitrap	QTOF	Orbitrap	QTOF	Orbitrap
4-aminoantipyrine ^a	7/1 ^b	6/5	7/7	7/3						
Acetaminophen	7/7	7/7 (7) ^c					2/2	2/2 (2)		
Atorvastatin			2/0	5/0						
Benzoylcegonine	7/7	7/7	7/7 (7)	7/7	1/0		2/2 (2)	2/2		
Bezafibrate		3/3	7/0	7/7			2/1 (1)	2/2		1/1
Caffeine ^{a,d}	7/7	6/6	7/1	7/7	4/3	5/5	2/2	2/2		
Carbamazepine	3/0	7/7	7/7	7/7			2/2	2/1	2/2	2/2
Ciprofloxacin	7/6 (6)	6/6	6/5 (5)	6/5			2/2 (2)	2/2		
Clarithromycin		1/0	5/5 (5)	6/6			2/1 (1)	2/1	2/1(1)	2/1
Cocaine	7/3 (3)	3/3	2/0	1/0			2/2 (2)	1/1		
Codeine	7/7 (7)	7/7	7/4 (4)	6/6			2/1 (1)	2/1	2/1(1)	1/1
Cotinine ^{a,d}	7/6		7/7		2/0		2/0			
Diclofenac		3/3	7/7	7/7			2/2	2/2	2/2	2/2
Enalapril	1/0									
Gabapentin ^a	5/0		6/6						2/0	
Gemfibrozil								2/0		
Irbesartan ^a	7/7	7/7	7/7	7/7	2/1	2/2	2/2	2/2	2/2	2/2
Ketoprofen	7/6 (6)	6/5	6/2	2/2			2/2			
Morphine	7/0	4/4	2/0	4/4			2/0	2/2	2/0	2/0
Moxifloxacin			2/0							
Naproxen	7/7	7/7 (3)	4/0				2/1	1/1 (1)		
Ofloxacin	7/3	7/7	7/7	7/7	1/0		2/2	2/2	2/2	2/2
Oxazepam ^{a,d}		3/3	7/7	5/5				1/1		1/0
Sulfamethoxazole	3/2 (2)	6/6	7/5	5/4			2/2 (2)	2/2	2/1(1)	2/2
Temazepam ^{a,d}	3/3 (3)	6/5	7/5	7/7				1/0		1/1
Trimethoprim	1/0		7/0	7/0			2/0		2/0	2/0
Valsartan ^a	7/7	7/7	7/7	7/7	2/0	3/3	2/2	2/2	1/1	2/2
Venlafaxine ^a	7/7 (6)	7/7	7/7	7/7		1/0	2/0	2/2	2/2	2/2
Total	21/16	23/18	25/18	21/18	6/2	4/4	20/16	20/16	12/9	12/10

^a: no reference standard for Orbitrap analysis

^b: number of compounds detected/identified

^c: number in brackets refers to the number of samples manually checked in raw data for missing fragment ions

^d: no reference standard for QTOF analysis

Comparison of results from Spain and Italy

Table 1 shows the number of compounds identified (according to identification criteria 2 and 4, from Section 2.5) and detected (according to detection criteria 1 and 3, from Section 2.5) in each sample. In general, more compounds were able to be identified (up to 18) or detected (up to 25) in the Spanish samples than the Italian. However, some compounds were only able to be detected and not identified in some samples, due to low analyte concentration. For example, ofloxacin in Castellón IWW was identified in only three samples by QTOF but detected in the all, whilst it was identified in all samples by Orbitrap. In addition, trimethoprim was only

detected in all Spanish and Italian EWW samples by both instruments. These results show that these compounds were likely present in all the detected samples but unable to be identified due to a combination of their low concentrations and the difference in MS/MS acquisition between the instruments.

There is little national difference, with just cotinine, gabapentin, oxazepam, temazepam and a metabolite of metamizole (4-aminoantipyrine), not identified in the Italian samples. This is not to say that these compounds are not in the samples, rather that their concentration is lower than that necessary for detection or identification. In fact, 4-aminoantipyrine was the only compound in the aforementioned list not detected in at least one of the Italian samples. Metamizole (also known as dipyrone) is a non-steroidal anti-inflammatory drug. In some countries such as Sweden, UK, Canada or USA, metamizole has been banned or restricted for human use due to its association with diseases like agranulocytosis [32]. However, in many other countries including Spain and Italy, it can be purchased without a medical prescription [33]. Metamizole is a widely consumed pharmaceutical in Spain, and some years ago, it was amongst the ten most consumed pharmaceuticals (considering the amount of prescriptions), according to the information published by the Spanish Ministry of Health [34]. In Italy, metamizole is also highly consumed but is not included in the list of the most consumed pharmaceuticals [35]. This apparent difference in consumption could reflect the difference in its detection in the samples of Spain and Italy.

The most commonly detected compounds mirrored those found in similar studies [12, 29, 30, 36]. Irbesartan, ofloxacin, sulfamethoxazole and valsartan were identified in IWW and EWW of both Spain and Italy. Caffeine, irbesartan, valsartan and venlafaxine were also found in a high enough concentration in surface water to be identified by at least one of the two instruments. Several previous studies have also shown that these compounds are not fully removed in WWTPs and can be present in surface waters [2, 5, 37, 38], which indicates that further investigation into their potential fate and harm to the environment is necessary.

Use of QTOF and LTQ-Orbitrap

There was little difference between the number of compounds able to be identified by QTOF and Orbitrap. However, there are some discrepancies between the numbers of compounds in each sample that are identified by each instrument. For example, naproxen is identified using the software in all seven Spanish IWW samples by QTOF whilst in only four by the Orbitrap. Alternatively, with Orbitrap, bezafibrate and morphine are identified in seven and four Spanish EWW samples, respectively, but in none by QTOF. In both cases, the compounds were detected and compared with standards. Figures 1 and 2 show the examples of naproxen and bezafibrate, respectively.

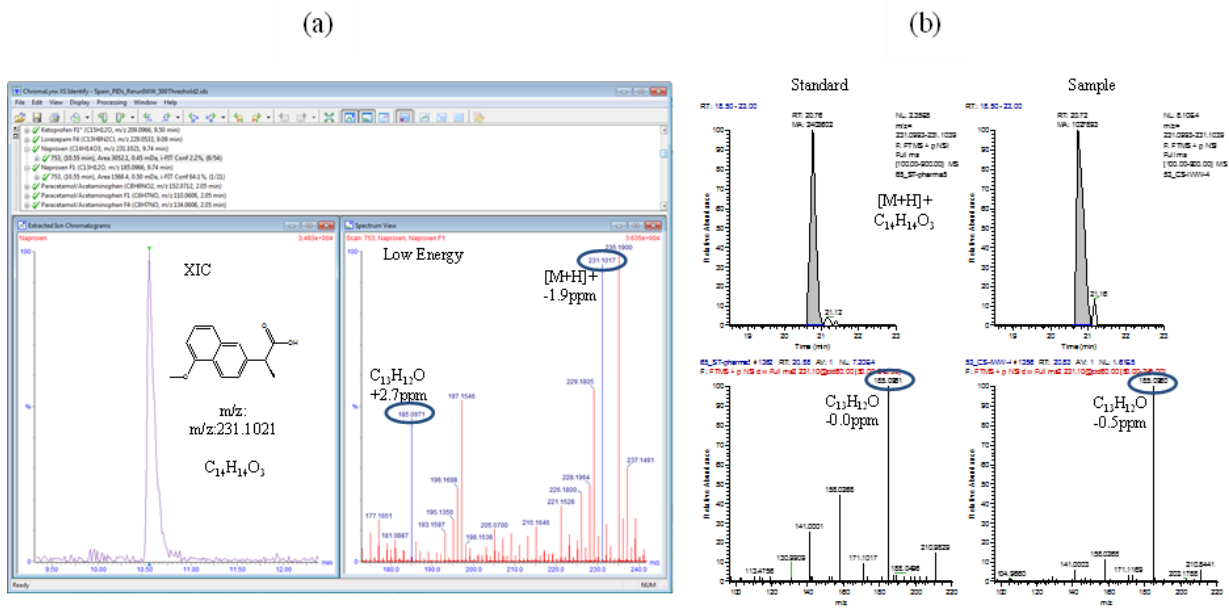


Fig. 1: (a): Identification of naproxen in an IWW sample by QTOF (Chromalynx); LE spectrum shows the parent and fragment of naproxen (blue at 231.1017 and 185.0971).

(b): Identification of naproxen in the same IWW sample by Orbitrap (raw data, using QualBrowser) based on the fragment of 185.0961, compared with the standard (left).

The QTOF data is from the single MS^E injection, while the Orbitrap data is from the second MS/MS injection.

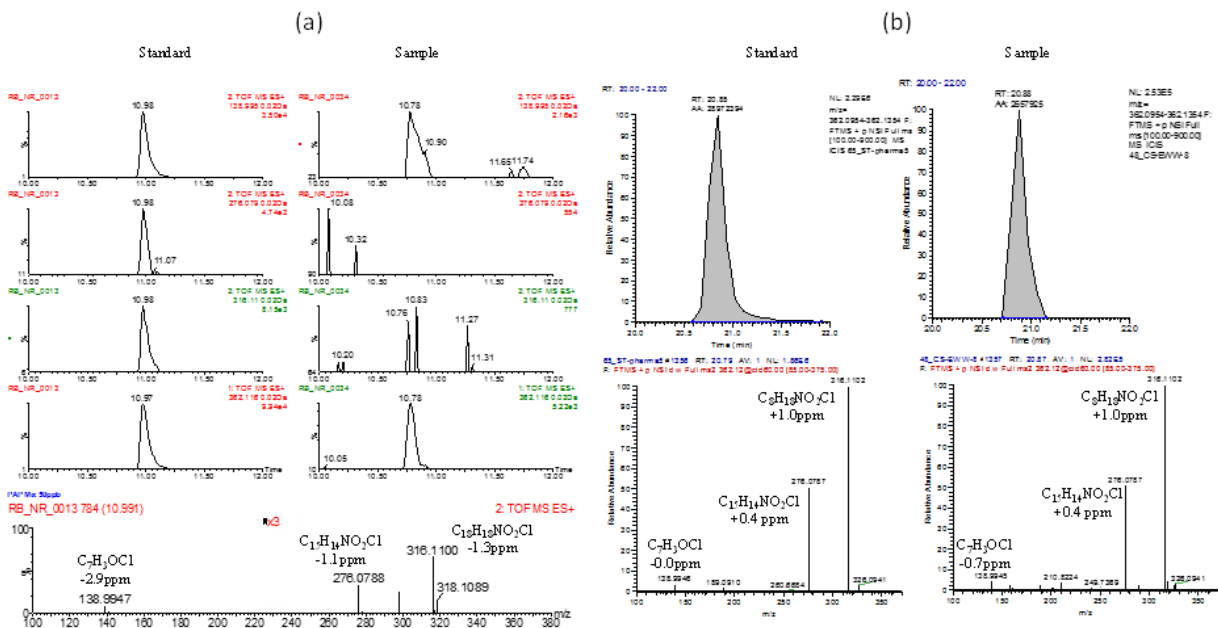


Fig. 2: (a) Detection of bezafibrate in an EWW sample by QTOF (raw data). The XICs (left) and related mass spectrum (bottom) of the standard show the fragments (316.1104, 276.0791 and 138.9951), and mass errors. The sample (right) shows none of the related fragments at a similar peak shape or retention time to the precursor (362.1159).

(b) Identification of bezafibrate in the same EWW sample (right) by Orbitrap (QualBrowser), compared with the standard (left). The fragment ions of 316.1102, 276.0787 and 138.9946 are clear in both, as is the chromatographic peak at ~20.85 minutes.

The QTOF data is from the single MS^E injection, while the Orbitrap data is from the second MS/MS injection.

Figure 1 provides an example of raw data investigation, touched upon in Section 3.1. In an IWW sample, the QTOF data was searched with ChromaLynx and able to identify naproxen, without having to refer to the raw data (Fig. 1a). The protonated molecule (231.1017) and primary fragment ion (185.0971) were seen in the LE and HE spectra within the thresholds of mass error for identification. To gain further confidence in this finding, the Orbitrap data was searched with TraceFinder. However, the software could not identify naproxen in this IWW sample. A subsequent search of the raw data by QualBrowser revealed its presence (Fig. 1b).

Figure 2 exemplifies the compatibility of the two HRMS instruments. ChromaLynx was unable to identify bezafibrate through the use of fragments, so the raw data was checked (Fig. 2a, right). Again, only the precursor was able to be found. A comparison with the standard is also made, showing the three fragments of 316.1105, 276.0788 and 138.9947 (Fig. 2a, left). These fragments were searched in the HE spectrum of the sample, but no suitable peaks were found (Fig. 2a, right). It was thought that bezafibrate was below the detection threshold for the QTOF, which has a known relationship between intensity and mass error (i.e. at low concentration, the mass error would be high), and therefore at a low concentration, the fragment ions would fall outside the given mass error threshold. Conversely, Orbitrap is able to maintain its mass accuracy, even at low concentrations. Hence, for the same sample, using Orbitrap, bezafibrate was able to be identified with all three fragments (Fig. 2b, right).

A further example of the use of these instruments was in the identification of benzoylecgonine in an IWW sample. The protonated ion was seen in the QTOF data with a mass error of 0.07 ppm, a retention time of Δ 0.01 min from that of the standard, and only one fragment with a mass error of 6.31 ppm—just outside the criteria used for identification (see Section 2.5). Although we were confident this was in fact benzoylecgonine, and did tentatively assign it as ‘identified’, it did not quite fulfil our set requirements. To appease our doubts, we also checked the Orbitrap data. Here, the protonated molecule and fragment ion were found within the threshold of 5 ppm and the retention time differed by 0.1 min from the standard. The isotopic pattern fit was above the 60 % threshold and the library search score of 50 %, above the threshold set. Therefore, full confidence in identification was gained for this compound. Although identification could also have been possible by manually scouring the raw data, as the fragment ion was only just outside the given threshold, it was thought better to compare with the Orbitrap data.

The mere fact that further information can be gleaned from these two instruments shows their complementariness. The European Directorate General for Health and Consumer Protection (SANCO) guidelines on analytical quality control and validation procedures for pesticide residue analysis in food and feed [39], which are also applied for environmental samples in the absence of specific guidelines for the environmental field, state that ‘confirmation is the combination of two or more analyses that are in agreement with each other’. The use of two HRMS instruments thus follows these SANCO guidelines, and as shown in Section 3.3 and Table 1, confirmation (i.e. identification) of numerous compounds is made. The examples given

thus demonstrate the complementary use of QTOF and Orbitrap in the suspect screening of PIDs in environmental and wastewaters.

Conclusions

In this paper, two LC-HRMS instruments (QTOF and LTQ-Orbitrap) used in different laboratories were used to screen for PIDs in samples of wastewater and surface water from Spain and Italy. A database of compounds frequently detected in water was used by both instruments, incorporating compounds of national significance. The samples were first investigated through the use of the manufacturer's software (ChromaLynx or TraceFinder), which allowed the detection of various compounds and in some cases the positive identification with fragment ions. Following this strategy, up to 18 compounds were identified with fragment ions across both instruments, whilst 28 compounds were at least detected (only the most abundant ion present) across all samples. In particular cases where identification was not automatic using software, the raw data was manually investigated. This was only necessary for a few compounds and occurred when the compound was detected but could not be identified. There was no significant difference between the countries, the metamizole metabolite, 4-aminoantipyrine, being the only compound found solely in Spanish waters.

The screening results obtained by QTOF MS and LTQ-Orbitrap MS were very similar, with the number of compounds detected/identified only differing by up to four across all samples. Although the objective of this work was not to compare the methods, rather the results in terms of detection and reliable identification, it is worth noting that the LTQ-Orbitrap was able to identify a few more compounds without having to manually search for fragment ions. On the whole, this paper shows that these powerful instruments, used in these configurations, are suitable for the detection and identification of numerous environmental contaminants in waste and environmental waters. The general agreement in data obtained will serve to validate the methodologies commonly used by our laboratories for qualitative screening of emerging contaminants in waters.

Acknowledgments

R. Bade, N. I. Rousis and E. Gracia-Lor acknowledge the European Union's International Training Network SEWPROF (Marie Curie-FP7-PEOPLE Grant no. 317205) for their Early Stage Researcher (R. Bade and N. I. Rousis) and Experienced Researcher (E. Gracia-Lor) contracts. The financial support from the Spanish Ministry of Economy and Competitiveness (Ref CTQ2012-36189) and Generalitat Valenciana, Spain (research group of excellence PROMETEO II/2014/023; ISIC 2012/016) is also acknowledged.

We would like to thank Thermo Scientific who provided the software TraceFinder and particularly Claudia Martins for technical support.

Conflict of interest

The authors declare that they have no competing interests.

References

1. Richardson SD (2012) Environmental mass spectrometry: emerging contaminants and current issues. *Anal Chem* 84:747–778. doi:10.1021/ac202903d
2. Li WC (2014) Occurrence, sources, and fate of pharmaceuticals in aquatic environment and soil. *Environ Pollut* 187:193–201. doi:10.1016/j.envpol.2014.01.015
3. Kümmerer K (2009) Antibiotics in the aquatic environment—a review—part I. *Chemosphere* 75:417–434. doi:10.1016/j.chemosphere.2008.11.086
4. Brozinski J, Lahti M, Meierjohann A, Oikari A, Kronberg L (2013) The anti-inflammatory drugs diclofenac, naproxen and ibuprofen are found in the bile of wild fish caught downstream of a wastewater treatment plant. *Environ Sci Technol* 47:342–348. doi:10.1021/es303013j
5. Luo Y, Guo W, Ngo HH, Nghiem LD, Hai FI, Zhang J et al (2014) A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci Total Environ* 473–474:619–641. doi:10.1016/j.scitotenv.2013.12.065
6. Pal R, Megharaj M, Kirkbride KP, Naidu R (2013) Illicit drugs and the environment—a review. *Sci Total Environ* 463–464:1079–1092. doi:10.1016/j.scitotenv.2012.05.086
7. Anumol T, Snyder SA (2015) Rapid analysis of trace organic compounds in water by automated online solid-phase extraction coupled to liquid chromatography–tandem mass spectrometry. *Talanta* 132:77–86. doi:10.1016/j.talanta.2014.08.011
8. Backe WJ, Field JA (2012) Is SPE necessary for environmental analysis? A quantitative comparison of matrix effects from large-volume injection and solid-phase extraction based methods. *Environ Sci Technol* 46:6750–6758. doi:10.1021/es300235z
9. Krauss M, Singer H, Hollender J (2010) LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal Bioanal Chem* 397:943–951. doi:10.1007/s00216-010-3608-
10. Hernández F, Ibáñez M, Bade R, Bijlsma L, Sancho JV (2014) Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry. *TrAC Trends Anal Chem* 63:140–157. doi:10.1016/j.trac.2014.08.003
11. Hernández F, Sancho JV, Ibáñez M, Abad E, Portolés T, Mattioli L (2012) Current use of high-resolution mass spectrometry in the environmental sciences. *Anal Bioanal Chem* 403:1251–1264. doi:10.1007/s00216-012-5844-7

12. Hernández F, Bijlsma L, Sancho JV, Díaz R, Ibáñez M (2011) Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography-quadrupole-time-of-flight-mass spectrometry. *Anal Chim Acta* 684:87–97. doi:10.1016/j.aca.2010.10.043
13. de Voogt P, Emke E, Helmus R, Panteliadis P, van Leerdam JA (2011) Determination of illicit drugs in the water cycle by LC-Orbitrap MS. In: Castiglioni S, Zuccato E, Fanelli R (eds.) *Illicit drugs environ. Occur. Anal. Fate Using Mass Spectrom.*, John Wiley & Sons, Ltd., pp. 87–114
14. Gómez MJ, Gómez-Ramos MM, Malato O, Mezcua M, Fernández-Alba AR (2010) Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography-quadrupole-time-of-flight mass spectrometry with an accurate-mass. *J Chromatogr A* 1217:7038–7054. doi:10.1016/j.chroma.2010.08.070
15. Gonzalez-Mariño I, Quintana JB, Rodriguez I, Gonzalez-Diez M, Cela R (2012) Screening and selective quantification of illicit drugs in wastewater by mixed-mode solid-phase extraction and quadrupole-time-of-flight liquid chromatography–mass spectrometry. *Anal Chem* 84:1708–1717. doi:10.1021/ac202989e
16. Ibáñez M, Guerrero C, Sancho JV, Hernández F (2009) Screening of antibiotics in surface and wastewater samples by ultra-high-pressure liquid chromatography coupled to hybrid quadrupole time-of-flight mass spectrometry. *J Chromatogr A* 1216:2529–2539. doi:10.1016/j.chroma.2009.01.073
17. Pinhancos R, Maass S, Ramanathan DM (2011) High-resolution mass spectrometry method for the detection, characterization and quantitation of pharmaceuticals in water. *J Mass Spectrom* 46:1175–1181. doi:10.1002/jms.2005
18. Wille K, Claessens M, Rappé K, Monteyne E, Janssen CR, De Brabander HF et al (2011) Rapid quantification of pharmaceuticals and pesticides in passive samplers using ultra high performance liquid chromatography coupled to high resolution mass spectrometry. *J Chromatogr A* 1218:9162–9173. doi:10.1016/j.chroma.2011.10.039
19. Kosma CI, Lambropoulou DA, Albanis TA (2014) Investigation of PPCPs in wastewater treatment plants in Greece: occurrence, removal and environmental risk assessment. *Sci Total Environ* 466–467:421–438. doi:10.1016/j.scitotenv.2013.07.044
20. SEWPROF ITN: a new paradigm in drug use and human health risk assessment: sewage profiling at the community level, (n.d.). <http://sewprof-itn.eu/>. Accessed 1 June 2015
21. Bijlsma L, Beltrán E, Boix C, Sancho JV, Hernández F (2014) Improvements in analytical methodology for the determination of frequently consumed illicit drugs in urban wastewater. *Anal Bioanal Chem* 406:4261–4272. doi:10.1007/s00216-014-7818-4

22. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP et al (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* 48:2097–2098. doi:10.1021/es5002105
23. Díaz R, Ibáñez M, Sancho JV, Hernández F (2011) Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 25:355–369. doi:10.1002/rcm.4860
24. Diaz R, Ibáñez M, Sancho JV, Hernández F (2013) Qualitative validation of a liquid chromatography-quadrupole-time of flight mass spectrometry screening method for organic pollutants in waters. *J Chromatogr A* 1276:47–57. doi:10.1016/j.chroma.2012.12.030
25. Hernández F, Ibáñez M, Portolés T, Cervera MI, Sancho JV, López FJ (2015) Advancing towards universal screening for organic pollutants in waters. *J Hazard Mater* 282:86–95. doi:10.1016/j.jhazmat.2014.08.006
26. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K et al (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* 45:703–714. doi:10.1002/jms.1777
27. Gracia-Lor E, Sancho JV, Hernandez F (2010) Simultaneous determination of acidic, neutral and basic pharmaceuticals in urban wastewater by ultra high-pressure liquid chromatography-tandem mass spectrometry. *J Chromatogr A* 1217:622–632. doi:10.1016/j.chroma.2009.11.090
28. Gracia-Lor E, Sancho JV, Hernandez F (2011) Multi-class determination of around 50 pharmaceuticals, including 26 antibiotics, in environmental and wastewater samples by ultra-high performance liquid chromatography-tandem mass spectrometry. *J Chromatogr A* 1218:2264–2275. doi:10.1016/j.chroma.2011.02.026
29. Zuccato E, Castiglioni S, Bagnati R, Melis M, Fanelli R (2010) Source, occurrence and fate of antibiotics in the Italian aquatic environment. *J Hazard Mater* 179:1042–1048. doi:10.1016/j.jhazmat.2010.03.110
30. Zuccato E, Castiglioni S, Fanelli R, Reitano G, Bagnati R, Chiabrando C et al (2006) Pharmaceuticals in the environment in Italy: causes, occurrence, effects and control. *Environ Sci Pollut Res Int* 13:15–21. doi:10.1065/espr2006.01.004
31. Al Aukidy M, Verlicchi P, Jelic A, Petrovic M, Barcelò D (2012) Monitoring release of pharmaceutical compounds: occurrence and environmental risk assessment of two WWTP effluents and their receiving bodies in the Po Valley, Italy. *Sci Total Environ* 438:15–25. doi:10.1016/j.scitotenv.2012.08.061
32. Gómez MJ, Sirtori C, Mezcuca M, Fernández-Alba AR, Agüera A (2008) Photodegradation study of three dipyrone metabolites in various water systems: identification and toxicity of their photodegradation products. *Water Res* 42:2698–2706. doi:10.1016/j.watres.2008.01.022

33. Metamizolo: una lunga storia tra luci ed ombre. Accessed November 2014 <http://www.farmacovigilanza.org/focus/200608/>
34. N. 2/2006 Prescription data: IT del Sistema Nacional de Salud Volumen 30, Subgrupos ATC y Principios activos de mayor consumo en el Sistema Nacional de Salud en 2005, (2006) 42–49
35. A.I. del Farmaco, L'uso dei Farmaci in Italia (2013) http://www.agenziafarmaco.gov.it/sites/default/files/Rapporto_OsMED_2013.pdf
36. Gracia-Lor E, Sancho JV, Serrano R, Hernandez F (2012) Occurrence and removal of pharmaceuticals in wastewater treatment plants at the Spanish Mediterranean area of Valencia. *Chemosphere* 87:453–462. doi:10.1016/j.chemosphere.2011.12.025
37. Ferrer I, Thurman EM (2012) Analysis of 100 pharmaceuticals and their degradates in water samples by liquid chromatography/quadrupole time-of-flight mass spectrometry. *J Chromatogr A* 1259:148–157. doi:10.1016/j.chroma.2012.03.059
38. Kasprzyk-Hordern B, Dinsdale RM, Guwy AJ (2007) Multi-residue method for the determination of basic/neutral pharmaceuticals and illicit drugs in surface water by solid-phase extraction and ultra performance liquid chromatography-positive electrospray ionisation tandem mass spectrometry. *J Chromatogr A* 1161:132–145. doi:10.1016/j.chroma.2007.05.074
39. Guidance document on analytical quality control and validation procedures for pesticides residues in food and feed, SANCO /12571/2013 (2013) http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/qualcontrol_en.pdf

Electronic Supplementary Material

The tables “List of Compounds” (Table S1), Compounds search in Italian samples by Orbitrap (Table S2) and Compounds search in Spanish samples by Orbitrap (Table S3) are too large for this thesis, but are available in the online edition of this article: doi:10.1007/s00216-015-9063-x

2.1 Reagents

2.1.1 Illicit Drugs (for QTOF)

In total, 11 illicit drugs and/or main metabolites were selected: amphetamine, methamphetamine, 3,4-methylenedioxymethamphetamine (MDMA, or ecstasy), cocaine, benzoylecgonine (BE), 11-nor-9- carboxy- Δ^9 -tetrahydrocannabinol (THC-COOH), heroin, morphine, 6-monoacetylmorphine (6-MAM), codeine and ketamine. Reference standards of these compounds were purchased from Cerilliant (Round Rock, TX, USA) and the National Measurement Institute (Pymble, Australia) as solutions in methanol, acetonitrile or as salt. HPLC-grade methanol (MeOH), HPLC-grade acetonitrile (ACN), ammonium acetate, formic acid (>98 %) and primary secondary amine (PSA, 40–60 μm) sorbent were acquired from Scharlau (Barcelona, Spain). HPLC-grade water was obtained by purifying demineralised water in a Milli-Q plus system from Millipore (Bedford, MA, USA). Standard stock solutions of each compound were prepared at 100 mg/L in MeOH or ACN. Intermediate solutions (10 mg/L) were prepared by diluting the stock solution ten times with MeOH. Infusion solutions of individual standards were prepared at a concentration of 1.5 mg/L in MeOH:water (50:50, v/v). Mixed working solution of 25 $\mu\text{g/L}$ containing all analytes was prepared from intermediate solutions by appropriate dilution with Milli-Q water.

2.1.2 Illicit Drugs (for Orbitrap)

Illicit drugs and metabolites analyzed were the following: cocaine and its metabolites benzoylecgonine, norbenzoylecgonine, norcocaine, and cocaethylene; amphetamine, methamphetamine, 3,4-methylenedioxyamphetamine (MDA), 3,4-methylenedioxymethamphetamine (MDMA, or ecstasy), 3,4-methylenedioxyethamphetamine (MDEA) and MBDB; ketamine, norketamine, dehydronorketamine; codeine, 6-aceyl codeine; oxycodone, hydrocodone; morphine, morphine-3 β -D-glucuronide, 6-monoacetylmorphine (a metabolite of heroin); 11-nor-9-carboxy- Δ^9 -tetrahydrocannabinol (11-nor-9- carboxy- Δ^9 -THC), the main metabolite of Δ^9 -tetrahydrocannabinol; and methadone and its main metabolite, 2-ethylidene-1,5- dimethyl-3,3-diphenylpyrrolidine (EDDP). Reference standards of all the substances were acquired from Cerilliant Corporation (Round Rock, Texas, USA). The standards, available as solutions in methanol or acetonitrile (1. 0.4 mg/mL or 0.1 mg/ mL), were diluted to 10 ng/ μL with methanol and stored at -20 °C in

the dark. Working solutions containing all the substances to be analyzed were prepared before each analytical run.

2.1.3 Pharmaceuticals (for QTOF)

The pharmaceutical standard mix was made from a combination of human and veterinary pharmaceuticals and/or main metabolites (in alphabetical order):

4-Acetylamino-Antipyrine, Acetaminophen, Alprazolam, Ampicilline, Atorvastatin, Azithromycin, Bezafibrate, Carbamazepine, Cefotaxime, Cephalexin, Chloramphenicol, Ciprofloxacin, Clarythromycin, Clenbuterol, Clindamycin, Cloxacillin, Diclofenac, Dicloxacillin, Enalapril, Enrofloxacin, Erythromycin A, Flumequine, Furazolidone, Gabapentin, Gemfibrozil, Ibuprofen, Irbesartan, Ketoprofen, Lincomycin, Lorazepam, Marbofloxacin, Moxifloxacin, Naproxen, Nalidixic acid, Norfloxacin, Ofloxacin, Olanzapine, Oxacillin, Oxolinic acid, Oxytetracycline, Pefloxacin, Penicillin G, Pipedimic acid, Piperacillin, Pravastatin, Roxythromycin, Salicylic acid, Spiramycin, Sulfadiazine, Sulfadoxine, Sulfamerazine, Sulfamethazine, Sulfamethoxazole, Tetracycline, Thiabendazole, Temazepam, Tiamulin, Trimethoprim, Valsartan, Venlafaxine. All standards were purchased from Across Organics (Geel, Belgium), Aventis Pharma (Madrid, Spain), Bayer Hispania (Barcelona, Spain), Cerilliant (Round Rock, TX, USA), Fluka (Buchs, Switzerland), Dr. Ehrenstorfer (Augsburg, Germany), Fort Dodge Veterinaria (Gerona, Spain), National Measurement Institute (Pymble, Australia), Riedel-de Haën (Seelze, Germany), Sigma Aldrich (St Louis, MO, USA), Vetoquinol Industrial (Madrid, Spain), and Witega (Berlin, Germany). All reference materials had purities higher than 98% (w/w), except for marbofloxacin and pefloxacin, which had purities higher than 93%.

The standard stock solutions of each compound ranged between 19 and 1200mg/L, prepared in MeOH or ACN. These compounds were separated into 6 separate intermediate working mixtures of 5mg/L. Prior to analysis, a working standard mixture of 50µg/L was made from the intermediate solutions by diluting with water.

2.1.4 Pharmaceuticals (for Orbitrap)

The following pharmaceuticals were included in the standard mix: Acetaminophen, Atenolol, Atorvastatin, Bezafibrate, Carbamazepine, Ciprofloxacin, Clarythromycin, Diazepam, Diclofenac, Enalapril, Erythromycin A, Furosemide, Gemfibrozil, Hydrochlorothiazide, Ibuprofen, Ketoprofen, Lincomycin, Naproxen, Ofloxacin, Omeprazole, Oxytetracycline, Ranitidine, Spiramycin, Sulfamethoxazole.

All reference standards were purchased from Sigma-Aldrich (Steinheim, Germany), ICN Biochemicals (Meckenheim, Germany) or Glaxo SmithKline (Philadelphia, USA); clarythromycin and omeprazole were obtained from pharmaceutical preparations for intravenous injection, respectively, from Abbott, Latina, Italy

(Klacid® i.v. 500 mg) and Malesci, Florence, Italy (Omeprazen®, 40 mg). All the reference standards were stored at 4 °C.

Standards were dissolved in methanol (except those for intravenous injection that were dissolved in saline) up to a concentration of 1 mg/mL and subsequently diluted to 10 ng/L (stock solutions). Three standard mixtures, containing all the pharmaceuticals to be analysed, were prepared before each analytical run by diluting stock solutions in methanol. The stock solutions of ciprofloxacin, ofloxacin, omeprazole, oxytetracycline and sulfamethoxazole were renewed monthly because of their limited stability. Purity of the stock solutions was checked before each analytical run by HPLC–MS–MS. All stock solutions were stored at –20 °C in the dark.

2.2.3 Scientific Article 4

**FACILITATING HIGH RESOLUTION MASS
SPECTROMETRY DATA PROCESSING FOR SCREENING
OF ENVIRONMENTAL WATER SAMPLES: AN
EVALUATION OF TWO DECONVOLUTION TOOLS**

Richard Bade, Ana Causanilles, Erik Emke, Lubertus Bijlsma, Juan V. Sancho, Felix Hernandez, Pim de Voogt

Submitted to Science of the Total Environment

Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools

Richard Bade^{a,1}, Ana Causanilles^b, Erik Emke^b, Lubertus Bijlsma^a, Juan V. Sancho^a, Felix Hernandez^a, Pim de Voogt^{b,c,*}

^a Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain.

^b KWR Watercycle Research Institute, Chemical Water Quality and Health, P.O. Box 1072, 3430 BB, Nieuwegein, The Netherlands

^c Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE, Amsterdam, The Netherlands

* Corresponding author: w.p.devoogt@uva.nl Tel.: +31 20 5256565; fax: +31 20 5257431

¹ Visiting researcher at KWR Watercycle Research Institute

Abstract

A screening approach was applied to influent and effluent wastewater samples. After injection in a LC-LTQ-Orbitrap, data analysis was performed using two deconvolution tools, MsXelerator (modules MPeaks and MS Compare) and Sieve 2.1. The outputs were searched incorporating an in-house database of more than 200 pharmaceuticals and illicit drugs or ChemSpider. This hidden target screening approach led to the detection of numerous compounds including the illicit drug cocaine and its metabolite benzoylecgonine and the pharmaceuticals carbamazepine, gemfibrozil and losartan. The compounds found using both approaches were combined, and isotopic pattern and retention time prediction were used to filter out false positives. The remaining potential positives were reanalysed in MS/MS mode and their product ions were compared with literature and/or mass spectral libraries. The inclusion of the chemical database ChemSpider led to the tentative identification of several metabolites, including paraxanthine, theobromine, theophylline and carboxylosartan, as well as the pharmaceutical phenazone. The first three of these compounds are isomers and they were subsequently distinguished based on their product ions and predicted retention times. This work has shown that the use deconvolution tools facilitates non-target screening and enables the identification of a higher number of compounds.

Keywords: Non-target screening, peak-picking, hidden target screening, software, high resolution mass spectrometry, aquatic environment

Introduction

The investigation of emerging contaminants has become prevalent in analytical environmental chemistry circles. The use of pharmaceuticals, personal care products and illicit drugs is increasing worldwide, due to the growing population and the rise in available products and the amount of these contaminants entering the aquatic environment is of concern (Fatta-Kassinos et al., 2011). There is no blanket removal process able to be undertaken by wastewater treatment plants (WWTPs) for all compounds, leading to poor removal rates and detection of many of these compounds in effluent wastewaters (EWW) and consequently in surface waters (Bijlsma et al., 2012; Luo et al., 2014; van der Aa et al., 2013).

High Resolution Mass Spectrometry (HRMS) instruments, such as Time of Flight (TOF) and Orbitrap have revolutionized the investigation of emerging contaminants in the aquatic environment due to their high sensitivity in full scan mode, their increased mass accuracy and the possibility to distinguish the isotopic pattern. HRMS instruments have the ability to screen for unknowns due to exact mass measurements and these are unique characteristics compared to other mass spectrometry instruments. Hybrid systems i.e. HRMS hyphenated to a quadrupole or linear ion trap (LTQ), such as the LTQ-Orbitrap, combines the tandem mass spectrometric capability associated with the LTQ with the high mass resolving power (up to 100,000 FWHM) and mass accuracy capability of the Orbitrap (de Voogt et al., 2011; Makarov and Scigelova, 2010). These hybrid configurations based on HRMS allow reliable interpretation of MS/MS spectra and are very valuable when dealing with complex environmental matrices, such as wastewater, where co-elution of analytes with matrix interferences can result in ambiguous peaks (Hogenboom et al., 2009). By utilising the ultra-high resolution capabilities, isobaric compounds can easily be differentiated (Hernández et al., 2012).

In the literature, three different approaches are described for the detection and/or identification of compounds: target, suspect/post-target and non-target (Aceña et al., 2015; Bletsou et al., 2015; Gago-Ferrero et al., 2015; Hernández et al., 2015b, 2014, 2005; Krauss et al., 2010; Leendert et al., 2015). Target methods are limited to a restricted number of compounds, for all of which reference standards must be obtained and, therefore, information on the occurrence of other unknown, relevant micropollutants may be missed. Suspect screening takes advantage of a database of “known” compounds, including molecular formulae, fragmentation and retention time, which can then be computationally correlated to spectral HRMS data to give potential positive compounds. As the concept of suspect screening implies that reference standards are not necessarily available, the tentative identification of potential positives needs to be confirmed by the use of reference standards (and MS/MS injections, if required) in a final step.

The third, non-target, approach is of increasing interest but notoriously difficult to undertake, as, strictly speaking, no a priori information is available (Krauss et al., 2010; Schymanski et al., 2014b; Zedda and Zwiener, 2012). Even with the help of automated peak-picking software, thousands of peaks can be detected in an individual sample (Hug et al., 2014; Kern et al., 2009). Consequently, subsequent steps must then be made to

reduce the number of peaks to a more manageable number, including molecular formula derivation, isotopic pattern, mass defect analysis and retention time prediction (Gago-Ferrero et al., 2015; Helbling et al., 2010; Kind and Fiehn, 2007). Further confidence in the “potential positives” remaining can be gained through the use of fragmentation in a subsequent MS/MS injection and comparison with *in silico* fragmentation and/or mass spectral libraries (Bletsou et al., 2015; Herrera-Lopez et al., 2014; Hug et al., 2014; Little et al., 2012), with the latter referred to as “hidden targets” (Letzel et al., 2015). In these situations, it is of prime importance and for ease of the analyst to have software capable of fulfilling most (if not all) of these steps automatically. Most manufacturers have software specific for their instrument and data, which can automatically extract analytes of interest from the raw data, to facilitate suspect screening approaches. However, despite the tremendous advances in software for metabolite/transformation product detection and further non-target work, sometimes not all required information is available in one platform, leading users to manufacturer-independent software, such as the Eawag open-source R-code packages *enviMass*, *enviPick*, *nontarget* and *RMassBank* (Schollée et al., 2015; Schymanski et al., 2014a) which can enable the incorporation of additional parameters, such as the steps outlined above. In spite of these problems, non-target screening is necessary to identify new or unknown relevant pollutants, which is why efforts need to be made in developing proper software and efficient identification tools.

This work portrays the combination of non-target data processing and hidden target searching of environmental water samples after injection in an LC-LTQ-Orbitrap. Two computational programs were utilized: *MsXelerator* (*MsMetrix*) and *Sieve 2.1* (Thermo Scientific). An in-house database of more than 200 pharmaceuticals, personal care products and illicit drugs was incorporated in both programs. Additionally, *Sieve 2.1* was used in combination with the *ChemSpider* search feature. The main objective was to demonstrate the utility and additional value of these software packages for screening. This led to the detection of numerous compounds across both programs. The compounds detected by both methods were then reinjected to obtain MS/MS fragmentation, leading to the tentative identification of 24 compounds. Ultimately, this work shows that the combined use of two deconvolution tools combined with two hidden target screening approaches provides more information than either one used individually.

2. Materials and Methods

2.1 Reagents

HPLC-grade methanol (MeOH), and formic acid (>98 % w/w) were purchased from Mallinckrodt Baker (Deventer, The Netherlands). The ultrapure water was obtained by purifying demineralized water in a Milli-Q system from Millipore (Bedford, MA, USA). SPE cartridges used were Oasis HLB 3 mL (60 mg) from Waters (Milford, MA, USA). Polytyrosine-1,3,6 standard used for mass axis calibration was purchased from Cs Bio Co. (Menlo Park, CA, USA). Mixed cellulose ester membrane filters (0.45 μm) were purchased from Whatman (Dassel, Germany).

2.2 Water samples and extraction procedure

Seven influent wastewater (IWW) and seven effluent wastewater (EWW) 24-hour composite samples were collected over seven consecutive days in March 2014. They were stored in high density polystyrene bottles, immediately centrifuged and stored in the dark at -20°C . Analyses were performed as soon as possible after collection in order to keep biotic or abiotic degradation to a minimum (Llorca et al. 2014).

A solid phase extraction (SPE) step was applied prior to analysis to pre-concentrate the samples. All samples were filtered through a mixed cellulose ester membrane filter ($0.45\ \mu\text{m}$). SPE was performed using Oasis HLB cartridges (60 mg). The water samples (EWW 100 mL, with IWW four times diluted (i.e. 25 mL sample diluted to 100mL by adding Milli-Q water)) were loaded onto the cartridges and reconstructed in 1 mL of 10:90 MeOH:H₂O after elution with MeOH (5 mL). A procedural blank was also made, following the steps above but using Milli-Q water. Analyses were performed by injecting 20 μL of the final extract (in triplicate) into the LC-LTQ FT Orbitrap. For further information on the SPE procedure, see (Hernández et al., 2015a).

2.3 Liquid Chromatography

The HPLC system, consisted of a Surveyor auto sampler model Plus and a Surveyor quaternary gradient HPLC-pump (Thermo Fisher Scientific, Breda, The Netherlands). Chromatographic separation of the compounds was made using an XBridge C18 column ($150\ \text{mm} \times 2.1\ \text{mm}$ I.D., particle size $3.5\ \mu\text{m}$) (Waters). The pre-column used was a $4.0\ \text{mm} \times 2.0\ \text{mm}$ I.D. Phenomenex Security Guard column (Bester, Amsterdam, the Netherlands). The analytical column and the guard column were maintained at a temperature of 21°C in a column thermostat. A gradient was used at a constant flow rate of $0.3\ \text{mL min}^{-1}$ using Milli-Q water (Solvent A) and MeOH (Solvent B) both with 0.05% formic acid. The percentage of organic modifier (B) was changed linearly as follows: 0 min, 5%; 40 min, 100%; 45 min, 100%; 47 min, 5%. Between consecutive runs, the analytical column was re-equilibrated for 5 min.

2.4 LTQ-FT Orbitrap mass spectrometry

An LTQ FT Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) was used. The LTQ part of this system was equipped with a Heated Ion Max Electrospray Ionization (HESI) probe and operated in the positive ion mode. The conditions were: source voltage 3.0 kV, heated capillary temperature 300°C , vaporizer temperature 350°C , capillary voltage 13 V and tube lens 70 V. Products ions were generated in the LTQ trap at a normalized collision energy setting of 35% and using an isolation width of 2 Da.

Full-scan accurate mass spectra (mass range from 50 to 1300 Da) were obtained at a mass resolution of 60,000. The total cycle time depends upon the resolution; at the selected resolution the total cycle time is 0.5 s. The instrument was initially set to operate in full-scan ('survey') mode with accurate mass measurements. When an ion exceeded a preset threshold, the instrument switched to product-ion scan mode in the ion trap part. Further details on instrument operating conditions can be found elsewhere (Bijlsma et al., 2013)

All data were acquired and processed using Xcalibur version 2.1 software. A second MS/MS injection was made by incorporating an inclusion list of masses (see Supporting Information (S.I.) Table S1 for list) with a retention window of ± 2 minutes and collision energy of 35%. Since MS/MS fragmentation was carried out in the ion trap, only nominal mass was measured.

Mass axis calibration was performed with every batch run just prior to starting the batch by using flow injection of a polytyrosine-1,3,6 solution ($[M+H]^+$ 182.01170/508.20783 and 997.39781) at a flow rate of $10 \mu\text{L min}^{-1}$.

2.5 Settings of Deconvolution tools

MsXelerator (MsMetrix)

MS Compare and MPeaks are modules within MsXelerator. MS Compare is specifically designed for comparing MS spectra, whereas the MPeaks module picks peaks, with the “keep largest C13 peaks only” and the “peak cluster” algorithms used to help discard some, the latter performing componentization which groups together all peaks (i.e. isotopes and adducts) arising from a single retention time. All samples were uploaded individually and later investigated as triplicates, corresponding to the three triplicate injections of each sample. Procedural blank samples were initially processed using the optimized software settings (see below) to subtract identical peaks from each wastewater sample.

The “peak picking” was carried out by MPeaks on each individual sample using the following parameters and values: Base peak width = 11 (arbitrary units); spike width = 5 scans; peak separation = 5 scans; peak threshold = 0.5% (vs. largest peak); smoothness threshold = 0.65%, signal/noise ratio = 20. The sensitivity value, which helps the user find more or less sensitive parameters for the previous three parameters, was set a relatively low, at 2 (out of a maximum setting of 6). The peaks picked using these parameters were further reduced by only keeping peaks relating to an $[M+H]^+$ charge state.

Using the second module, MS Compare, all samples were subjected to the following LC/MS settings (in accurate mass mode) of the module for peak picking across multiple samples: No baseline correction; FWHM (scans) = 3 (min) - 40 (max); min peak height = 10,000 counts; delete spikes; m/z range: 100-650; Max. shift between peaks for grouping: 20 scans; Time window for XIC: 0.25min; Mass accuracy: 10 ppm.

Sieve 2.1 (Thermo Scientific)

The “control compare trend” feature of Sieve 2.1 (Thermo Scientific) was used with the following parameters: peak intensity threshold = 250,000 (62,500 for IWW); m/z range = 100-650; m/z width = 10ppm; retention time range = 3-40 minutes; maximum number of frames = 5,000; frame time width = 1.00 minute; align bypass = true. For the hidden target screening, either the database used in the analyses of MPeaks and MS Compare or ChemSpider was incorporated.

2.6 General workflow

The general workflow followed in this work (pictorialized in Figure 1) falls into the “hidden target” area of non-target screening, hypothesized by Letzel et al. (Letzel et al., 2015), wherein non-target techniques (i.e. peak picking) are originally applied, but a database (i.e. in-house database or ChemSpider) is used for identification.

All samples were injected in triplicate and the data were processed with the different software packages of MsXelerator or Sieve 2.1. Only peaks in all three injections were further investigated. This resulted in a list of chromatographic peaks, based on the accurate masses of their protonated molecules. To gain a list of potential positives, two hidden target identification methods were used: 1) an in-house database, containing more than 200 parent compounds and metabolites and the online database ChemSpider. False positives were manually removed after investigating the isotopic pattern (for the characteristic patterns of sulfur- and chlorine-containing species) and retention time prediction. A final “target list” was investigated by reinjecting the samples in MS/MS mode, to get product ions. Fragmentation was then compared with online databases and literature, which allowed the tentative identification of several compounds.

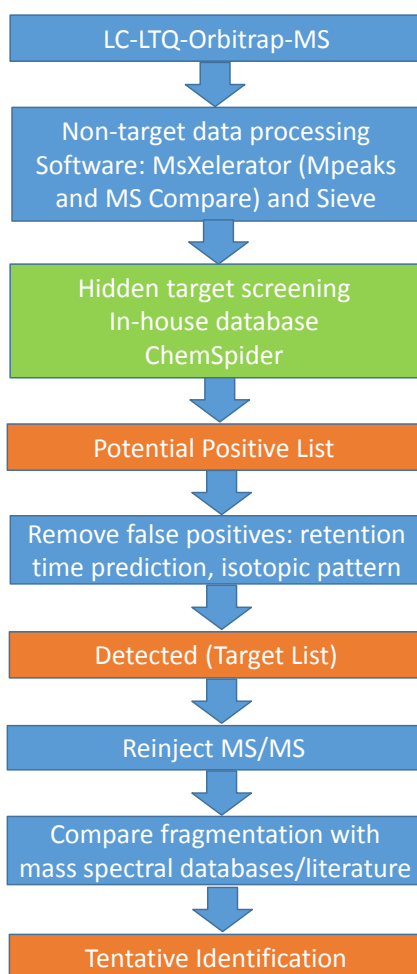


Figure 1: Workflow for screenings using the deconvolution tools MsXelerator and Sieve. All orange levels represent specific identification confidence levels

MPeaks and MS Compare are modules within MsXerator. Procedural blank samples were initially processed using the optimized software settings (Section 2.5) to subtract identical peaks from each wastewater sample. MS Compare is specifically designed for comparing MS spectra, whereas the MPeaks module picks peaks, with the “keep largest C13 peaks only” and the “peak cluster” algorithms used to help discard some, the latter performing componentization which groups together all peaks (i.e. isotopes and adducts) arising from a single retention time.

Sieve 2.1 combines the power of the two modules from MsXerator as described above. After an initial peak-picking process using the settings described in Section 2.5, it compares MS spectra of the procedural blank samples and the studied wastewater samples. Only compounds with an m/z between 150 and 500, and only protonated molecules ($[M+H]^+$) were considered. The list of potential positives were then search by either the in-house database or ChemSpider, which were incorporated in the software. The results of the ChemSpider search were exported into Microsoft Excel, and positive “hits” were considered based on their mass error (< 2 ppm), and if the compound commercially existed i.e. hits which only represented chemical formula were excluded.

3. Results and Discussion

In this study, in order to show the progression through confidence levels of identification, the terminology proposed in the literature by Hernández et al. (Hernández et al., 2015) and Schymanski et al. (Schymanski et al., 2014) were followed. It must be noted that potential positives and detected compounds, differentiated in this work, would both be level 3 tentative candidate in the terminology of Schymanski et al. The final, tentatively identified compounds are of a higher confidence level (level 2a). However, in order to have total confirmation (level 1), reference standards are necessary. As no reference standards were utilized in this work, this level could not be attained.

3.1 Optimization of deconvolution tools

All samples were injected and processed in triplicate, which were compared together, with only peaks in all three injections being further investigated. Procedural blank samples were processed first, to subtract identical peaks from each subsequent IWW and EWW sample.

The m/z range of MS Compare was made quite narrow as the compounds of interest in this study and in the in-house database (small pharmaceutical/drug molecules) would be within that range. The retention time range was reduced just to 3-40 mins to reduce the likelihood of erroneously detecting species that elute very early and late due to the high/low ratio of organic modifier, with the vast majority of all peaks in the total ion chromatogram falling within this range. In spite of the known mass accuracy capability of the Orbitrap, the mass accuracy was set at 10ppm to ensure that no compound would be missed. After this processing, a list of

masses common within each triplicate set was made, with compounds being detected using the same peak peaking parameters and database used in the final step of the MPeaks analysis.

Sieve 2.1 used the same m/z range, m/z width and retention time range as MS Compare for better ease of results comparison. The peak intensity threshold was originally set quite high for both IWW and EWW samples, but it was later found that IWW gave fewer peaks, possibly due to the complexity of these samples and stronger matrix effects, mostly leading to ionization suppression. The threshold was thus reduced to one quarter to account for this. The maximum number of components was raised to 10,000 to ensure that no compounds would be missed, leading to more than 5,000 components being detected in the IWW and EWW samples (Table 1). These were reduced by including compounds with only a m/z 150-500 and $[M+H]^+ = 1$. The in-house database used by the previous two modules within the “Accurate Mass Identification Parameters” of Sieve was then used to gain a list of potential positives.

The ChemSpider database (with 10ppm mass accuracy threshold) was also used within Sieve and was begun after the initial component optimization was completed (Figure 2). The threshold was made quite high, for optimal “hidden target” analysis, where the detected peaks should correspond to compounds which are commonly and/or highly used. The peak lists of both IWW and EWW results with all data pertaining to mass error, m/z and intensity were exported into Microsoft Excel. From these lists, several thresholds were set and outlying peaks removed: only compounds between m/z 150-500; only $[M+H]^+$; mass error under 2.0ppm; all “hits” just representing a chemical structure, rather than a generic/known name. This final step is rather pragmatic but makes for a more optimal non-target screening, where the remaining compounds should be the more common and/or highly used, as emphasized by having a high intensity threshold. However, this could lead to some less intense peaks being missed and not noted as a possible emerging contaminant in the environment.

Table 1: Number of components left after each step of the Sieve hidden target identification approaches.

Step	IWW components	EWW components
1.	6690	5091
2.	5158	3528
3.	2014	2175
4.	18	16
NON-TARGET SCREENING		
Number of distinct m/z (total number of compounds)		
5.	239 (437)	441 (677)
6.	166 (362)	308 (543)
7.	100 (150)	64 (108)
Final	18	8

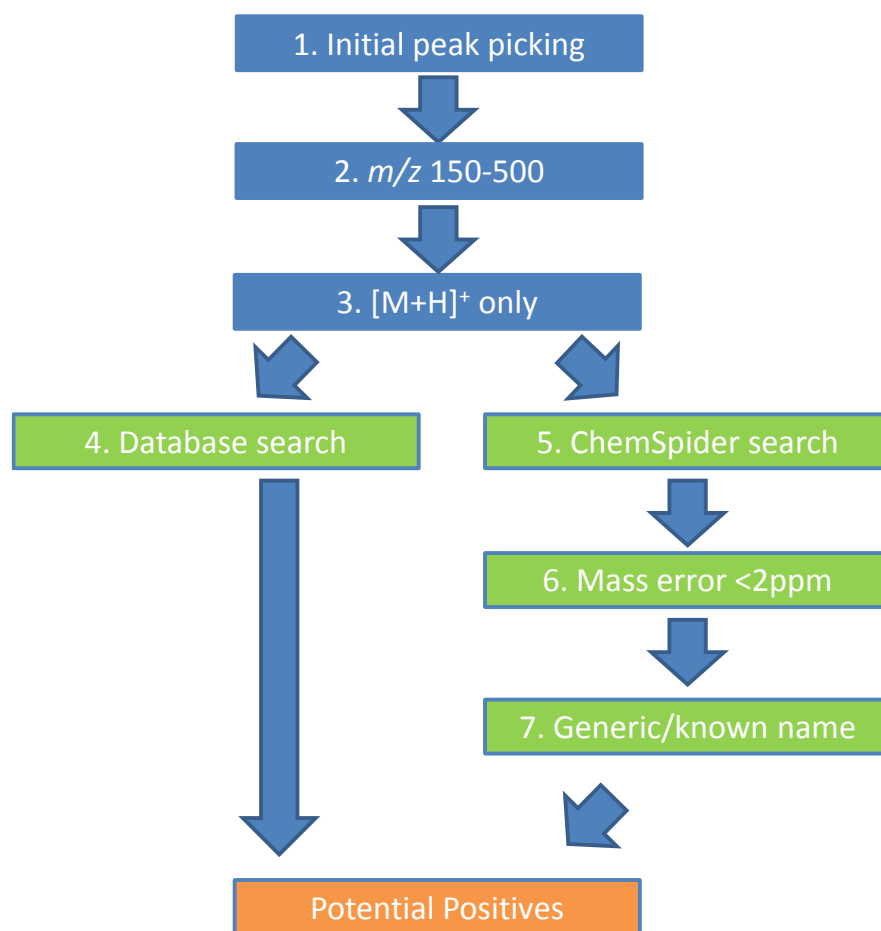


Figure 2: Sieve workflow for the two hidden target identification approaches.

3.2 Identification with in-house database

Both programs incorporated an in-house database of more than 200 pharmaceuticals, illicit drugs and metabolites (Table S3) to get a list of potential positives. All samples were first processed with MsXelerator (modules MPeaks and MSCompare) and Sieve 2.1 using the parameters outlined in Section 2.5. Table 2 shows the compounds detected by each.

There was very little difference between the compounds found with Sieve and MS Compare, while MPeaks detected somewhat fewer compounds. This could be due to their apparent uses: MPeaks is for pick-peaking, MS Compare for comparing samples, while Sieve does both, resulting in the latter two have more similar results. The fact that all compounds detected by MPeaks were also found with MS Compare leads to the preferential use of the latter module for screening. However, by optimizing the peak-peaking parameters of MPeaks, specifically the sensitivity value, this module could also be of future use in suspect and/or non-target screening.

Two methods were used to remove potential false positive peaks: isotopic pattern (for chlorine- and sulfur-containing species) and retention time prediction. Only three of the above compounds (losartan, sulfamethoxazole and temazepam) had a chlorine or sulfur atom, giving rise to a characteristic isotopic pattern.

Extracted ion chromatograms were extracted from the initial full-scan data of the Orbitrap and investigated manually. Both sulfamethoxazole and losartan showed the characteristic isotopic pattern, while temazepam did not. Temazepam was thus considered as a false positive and removed from further investigation.

Table 2: All compounds detected (in at least one sample) by each program following suspect screening.

Compound	IWW			EWW		
	Sieve	MSCompare	Mpeaks	Sieve	MSCompare	Mpeaks
4-acetylamino antipyrine	Not detected	Detected	Detected	Detected	Detected	Detected
4-formylamino-antipyrine	Not detected	Not detected	Not detected	Detected	Detected	Not detected
Acetaminophen	Detected	Detected	Detected	Detected	Detected	Detected
Benzocaine	Detected	Detected	Detected	Not detected	Detected	Not detected
Benzoylecgonine	Detected	Detected	Detected	Detected	Detected	Detected
Caffeine	Detected	Detected	Detected	Detected	Detected	Detected
Carbamazepine	Not detected	Not detected	Not detected	Detected	Detected	Detected
Cocaine	Detected	Detected	Detected	Not detected	Not detected	Not detected
Cotinine	Detected	Detected	Not detected	Detected	Detected	Not detected
Gemfibrozil	Detected	Not detected	Not detected	Not detected	Not detected	Not detected
Ibuprofen	Not detected	Not detected	Not detected	Detected	Detected	Detected
Irbesartan	Detected	Detected	Not detected	Detected	Detected	Not detected
Ketoprofen	Detected	Detected	Not detected	Detected	Detected	Not detected
Lidocaine	Detected	Detected	Detected	Not detected	Not detected	Not detected
Lincomycin	Detected	Detected	Detected	Detected	Detected	Detected
Losartan	Detected	Detected	Detected	Detected	Detected	Detected
Metoprolol	Detected	Detected	Detected	Not detected	Not detected	Not detected
Naproxen	Detected	Not detected	Not detected	Detected	Detected	Detected
Phenacetin	Not detected	Not detected	Not detected	Detected	Detected	Detected
Phenytoin	Detected	Not detected	Not detected	Not detected	Detected	Not detected
Salbutamol	Not detected	Detected	Detected	Not detected	Detected	Detected
Sulfamethoxazole	Detected	Detected	Detected	Detected	Detected	Detected
Temazepam	Not detected	Detected	Not detected	Not detected	Not detected	Not detected
Trimethoprim	Detected	Detected	Detected	Not detected	Detected	Not detected
Valsartan	Detected	Detected	Not detected	Detected	Detected	Detected

	Detected by the program
	Not detected by the program

A retention time predictor was made, based on artificial neural networks, as in (Miller et al., 2013; Munro et al., 2015) and in our previous work (Bade et al., 2015a). A retention time window of $\pm 11\%$ of total run time was used to find compounds to focus on, based on the window used in our previous work. Of the 25 potential positives investigated, four were removed using this method (benzocaine, ibuprofen, lincomycin and salbutamol) with predicted retention times between 11.5-16.8 minutes (24-36% of the total run time) away from the experimental times. While only four compounds were removed using this technique, it does simplify the identification process, and provides greater confidence in the compounds remaining.

3.3 Identification with ChemSpider

To make a more comprehensive analysis of the samples, an investigation was made using the ChemSpider database search feature of Sieve (Figure 2). The introduction of ChemSpider, while removing many components, had the added complication of isobaric and isomeric compounds, with most distinct m/z values having more than one compound associated, as seen in step 5 of Table 1. To further refine this list, the mass error was limited to 2ppm (step 6) and all compounds having a formula-only entry were deleted, leaving just compounds with generic names (step 7), leaving up to 100 components in the samples. The literature was then searched to determine whether or not their detection in wastewater could be expected, leading to approximately 30 components and up to 34 isomeric/isobaric compounds in the samples. The literature search was made using the Scopus database, and search terms were the generic name of interest, “HRMS”, “LC”, “environment” and “water”. If there were no suitable papers concerning the generic name of interest, the compound was removed from further investigation. To determine which of the isomeric/isobaric compounds the compound within the sample was, the molecular formula was manually searched on ChemSpider, with the compound having the highest number of references deemed to be the compound of interest. Step 7 and the literature search, while pragmatic, were employed to ensure that the compounds detected were those of high consumption/prescription and could therefore be more easily identified in the later in silico fragmentation comparison. Finally, eighteen (including three isomers) and eight compounds were finally deemed as potential positives using this non-target approach for IWW and EWW samples respectively (Table S4).

It is worth noting that by using this approach, most of the same compounds were found as with the in-house database (Table 2 and S4). With such great similarities between the set of potential positive compounds, only 2-hydroxy carbamazepine, desvenlafaxine, adenosine, albendazole, phenazone and the three isomers theophylline, paraxanthine and theobromine required further investigation. Albendazole was the only compound that required an investigation of isotopic pattern as it contains one sulfur atom, which was inconsistent with the mass spectrum, leading to its removal as a false positive. The remaining seven compounds were subjected to retention time prediction based on the time given by Sieve 2.1, and all were found within the set $\pm 11\%$ of total run time retention time window.

3.4 Tentative identification

Table 3: All compounds tentatively identified, together with retention time and fragment ions

Compound	m/z	RT	Fragment ions			IWW	EWW
4-acetylaminoantipyrine ¹	246.1234	9.43	228.1	204.1		X	X
4-formylaminoantipyrine ²	232.1086	9.29	214.1	204.1		X	X
Acetaminophen ³	152.0706	6.09	110.1	134.0		X	X
Adenosine ²	268.1035	3.58	136.0				X
Benzoylecgonine ²	290.1385	12.22	168.1			X	X
Caffeine ¹	195.0876	10.4	138.1			X	X
Carbamazepine ³	237.1022	22.56	194.1	152.9		X	X
Carboxylosartan ¹	437.1480	26.94	207.1	235.1	365.3	X	X
Cocaine ²	304.1543	13.84	182.2			X	
Ketoprofen ³	255.1014	16.78	237.1	209.1		X	X
Lidocaine ¹	235.1807	10.12	86.1			X	
Losartan ¹	423.1695	25.58	405.0	207.2	377.2	X	X
Metoprolol ²	268.1908	13.71	218.1	191.1	159.1	X	
Naproxen ³	231.1016	27.24	185.1			X	X
Paraxanthine ²	181.0721	7.98	124.1			X	X
Phenacetin ²	180.1030	17.22	138.1	110.0		X	X
Phenazone ²	189.1022	12.15	161.2	146.1	131.1	X	X
Sulfamethoxazole ²	254.0594	12.41	235.8	188.1	156.1	X	X
Theobromine ^{2,4}	181.0721	6.28	163.1	137.1	138.1	X	X
Theophylline ^{2,4}	181.0721	8.37	124.1			X	X
Trimethoprim ²	291.1454	9.91	230.2	123.2	261.1	X	X
Valsartan ²	436.2341	28.96	335.1	265.2	155.1	X	X

The parent ions were recorded at accurate mass (full-scan mode) while the fragment ions were recorded as part of a product ion scan in the ion trap part with nominal mass measurement.

¹ Information on fragment ions from Hernández *et al* (Hernández et al., 2015a)

² Information on fragment ions from MassBank (Horai et al., 2010)

³ Information on fragment ions from Bade *et al* (Bade et al., 2015c)

⁴ Information on fragment ions from Gómez *et al* (Gómez et al., 2010)

The potential positives found using both hidden target screening approaches were combined, less those removed in previous steps, leaving 28 compounds to investigate (Table S1). These compounds were added to a target list and several IWW and EWW samples were reinjected to see if fragment ions from these compounds

could give further confidence to their identification. Metfusion and MassBank were used to help provide further confidence to the fragment ions. As has been mentioned in previous suspect and non-target studies (Agüera et al., 2013; Herrera-Lopez et al., 2014; Zedda and Zwiener, 2012), the use and improvement of mass spectral databases, such as MassBank, is extremely important in the tentative identification of compounds for which standards are unavailable. In the end, 22 compounds were able to be tentatively identified (Table 3) with at least one fragment ion, while the other six were removed as false positives due to having incorrect fragment ions.

One interesting finding was the detection of three isomers (paraxanthine, theobromine and theophylline). These three isomers are all metabolites of caffeine, accounting for 80%, 11% and 4% of total metabolism, respectively (Miners and Birkett, 1996). Conventionally, isomeric compounds, separated chromatographically, would be distinguished by retention time. However, as no standards were available, the best way to order the peaks was with retention time prediction. The approach outlined in Section 3.3 predicted retention times of 8.48 min, 9.34 min and 9.41 min for theobromine, paraxanthine and theophylline, respectively. While these times are 1-2 minutes from the experimental retention time, they do provide an idea for the order of the isomers. To give more confidence to this information, the fragment ions were checked. As seen in Figure 3, the peak at 6.27 min had fragment ions of m/z 163 and 138 while the peaks at 7.98 and 8.37 both had one major peak of m/z 124. These fragment ions were checked and compared with MassBank and the literature (Bianco et al., 2009; Gómez et al., 2010; Horai et al., 2010).. Theobromine was found to have fragment ions of m/z 163 and 138, while both paraxanthine and theophylline were found to have a main fragment ion of m/z 124. The losses leading to each fragment ion is defined in Figure 3. To differentiate the latter two, the initial retention time predictions led to paraxanthine being the larger peak at 7.98 min and theophylline the small peak at 8.37 min.

While 22 compounds were tentatively identified using the workflow outlined throughout this paper, it must be noted that even incorporating the false positive removal strategies of retention time prediction and isotopic pattern as well as fragment ions, the final confirmation of the identity of compounds requires the use of reference standards. Nevertheless, the addition of advanced deconvolution tools (MsXerator and Sieve) to the HRMS data of the Orbitrap has been shown to be of great value, and the results show how far one can go without the need to purchase reference standards. The information obtained with this strategy circumvents the cost and problems associated with the storage and expiry dates of standards in the laboratories, as the purchase can be directed only towards those compounds that have been previously tentatively identified in the samples.

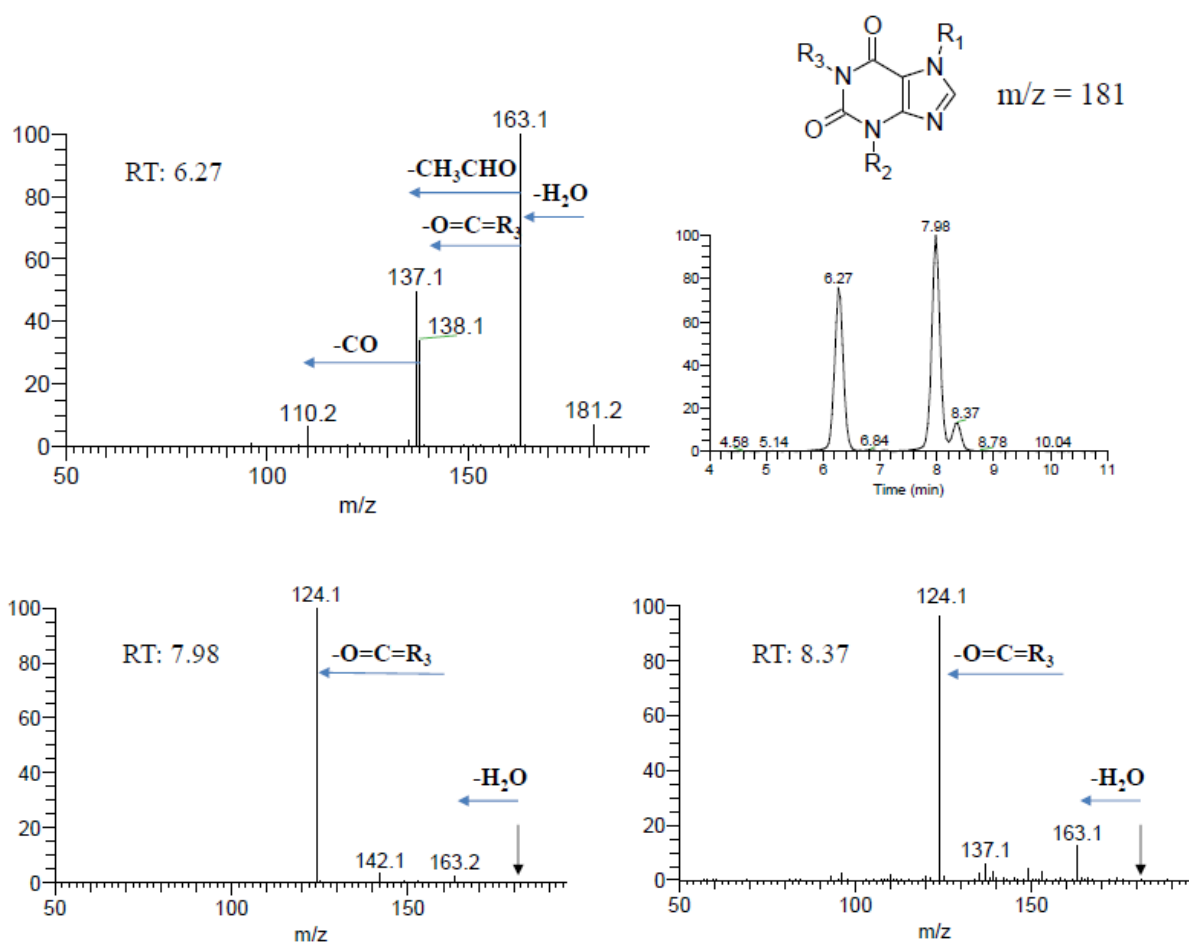


Figure 3: Tentative identification of theobromine (top left), paraxanthine (bottom left) and theophylline (bottom right), with chromatographic peaks (top right). The generic structure has been shown in the top right corner, where R₁, R₂ and R₃ differ for the metabolites as follows: theobromine: R₁ = CH₃, R₂ = CH₃ and R₃ = H; theophylline: R₁ = H, R₂ = CH₃ and R₃ = CH₃; paraxanthine: R₁ = CH₃, R₂ = H and R₃ = CH₃

Conclusion

This work has shown that, following initial unbiased, non-target oriented deconvolution using two tools (MsXerator and Sieve), allowed relevant peaks of interest to be attained. The complementary use of an in-house database or ChemSpider facilitated detection and enabled the identification of more compounds than using just one of these databases.

The combination of deconvolution tools and high resolution mass spectrometry, without the use of any reference standards, has enabled 22 compounds to be tentatively identified in environmental water samples. The majority of compounds that were identified in wastewater samples were pharmaceuticals, including the metabolites 4-formylamino antipyrine, 4-acetylamino antipyrine, theobromine, theophylline, paraxanthine and carboxylosartan.

It is worth noting that the two hidden target approaches primarily found the same compounds, with some exceptions. Furthermore, when applying small databases it is often easier to analyse the raw data directly. Whereas, when a much larger database is incorporated, these software tools will facilitate searching as well as reducing processing time. With further improvements to these computational programs non-target analysis will become more enticing and easier for laboratories to use in everyday screening methods.

Acknowledgements

Richard Bade and Ana Causanilles acknowledge the European Union for their Early Stage Researcher (ESR) contracts as part of the EU-International Training Network SEWPROF (Marie Curie- PEOPLE Grant #317205)

Part of this work was supported by the COST Action ES1307 “SCORE - Sewage biomarker analysis for community health assessment”.

The financial support of Generalitat Valenciana (Prometeo II 2014/023) and of the Spanish Ministry of Economy and Competitiveness (Project ref CTQ2015-65603) is also acknowledged by the authors of University Jaume I.

References

- Aceña, J., Stampachiachiere, S., Pérez, S., Barceló, D., 2015. Advances in liquid chromatography–high-resolution mass spectrometry for quantitative and qualitative environmental analysis. *Anal. Bioanal. Chem.* 6289–6299. doi:10.1007/s00216-015-8852-6
- Agüera, A., Martínez Bueno, M.J., Fernández-Alba, A.R., 2013. New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters. *Environ. Sci. Pollut. Res. Int.* 20, 3496–515. doi:10.1007/s11356-013-1586-0
- Bade, R., Bijlsma, L., Miller, T.H., Barron, L.P., Sancho, J.V., Hernández, F., 2015a. 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.* 538, 934–941. doi:10.1016/j.scitotenv.2015.08.078
- Bade, R., Rousis, N.I., Bijlsma, L., Gracia-Lor, E., Castiglioni, S., Sancho, J. V., Hernandez, F., 2015c. Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS. *Anal. Bioanal. Chem.* 407, 8979–8988. doi:10.1007/s00216-015-9063-x
- Bianco, G., Abate, S., Labella, C., Cataldi, T.R.I., 2009. Identification and fragmentation pathways of caffeine metabolites in urine samples via liquid chromatography with positive electrospray ionization coupled to a hybrid quadrupole linear ion trap (LTQ) and Fourier transform ion cyclotron resonance mass spec. *Rapid Commun. Mass Spectrom.* 23, 1065–1074. doi:10.1002/rcm.3969
- Bijlsma, L., Emke, E., Hernandez, F., de Voogt, P., 2012. Investigation of drugs of abuse and relevant metabolites in Dutch sewage water by liquid chromatography coupled to high resolution mass spectrometry. *Chemosphere* 89, 1399–1406. doi:10.1016/j.chemosphere.2012.05.110
- Bijlsma, L., Emke, E., Hernández, F., de Voogt, P., Hernandez, F., 2013. Performance of the linear ion trap Orbitrap mass analyzer for qualitative and quantitative analysis of drugs of abuse and relevant metabolites in sewage water. *Anal. Chim. Acta* 768, 102–110. doi:10.1016/j.aca.2013.01.010
- Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. *Trends Anal. Chem.* 66, 32–44.
- de Voogt, P., Emke, E., Helmus, R., Panteliadis, P., van Leerdam, J.A., 2011. Determination of illicit drugs in the water cycle by LC-Orbitrap MS, in: Castiglioni, S., Zuccato, E., Fanelli, R. (Eds.), *Illicit Drugs in the Environment: Occurrence, Analysis, and Fate Using Mass Spectrometry*. John Wiley & Sons, Ltd., pp. 87–114.

- Fatta-Kassinos, D., Meric, S., Nikolaou, A., 2011. Pharmaceutical residues in environmental waters and wastewater: Current state of knowledge and future research. *Anal. Bioanal. Chem.* 399, 251–275. doi:10.1007/s00216-010-4300-9
- Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* 49, 12333–12341. doi:10.1021/acs.est.5b03454
- Gerlich, M., Neumann, S., 2013. MetFusion: integration of compound identification strategies. *J. Mass Spectrom.* 48, 291–8. doi:10.1002/jms.3123
- Gómez, M.J., Gómez-Ramos, M.M., Malato, O., Mezcua, M., Fernández-Alba, A.R., 2010. Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography-quadrupole-time-of-flight mass spectrometry with an accurate-mass. *J. Chromatogr. A* 1217, 7038–54. doi:10.1016/j.chroma.2010.08.070
- Helbling, D.E., Hollender, J., Kohler, H.-P.E., Singer, H., Fenner, K., 2010. High-throughput identification of microbial transformation products of organic micropollutants. *Environ. Sci. Technol.* 44, 6621–7. doi:10.1021/es100970m
- Hernández, F., Ibáñez, M., Bade, R., Bijlsma, L., Sancho, J.V., 2014. Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry. *TrAC Trends Anal. Chem.* 63, 140–157. doi:10.1016/j.trac.2014.08.003
- Hernández, F., Ibáñez, M., Botero-Coy, A.-M., Bade, R., Bustos-López, M.C., Rincón, J., Moncayo, A., Bijlsma, L., 2015a. LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia. *Anal. Bioanal. Chem.* 407, 6405–6416. doi:10.1007/s00216-015-8796-x
- Hernández, F., Ibáñez, M., Portolés, T., Cervera, M.I., Sancho, J. V., López, F.J., 2015b. Advancing towards universal screening for organic pollutants in waters. *J. Hazard. Mater.* 282, 86–95. doi:10.1016/j.jhazmat.2014.08.006
- Hernández, F., Pozo, Ó.J., Sancho, J. V., López, F.J., Marín, J.M., Ibáñez, M., 2005. Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC–MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. *TrAC Trends Anal. Chem.* 24, 596–612. doi:10.1016/j.trac.2005.04.007
- Hernández, F., Sancho, J. V., Ibáñez, M., Abad, E., Portolés, T., Mattioli, L., 2012. Current use of high-resolution mass spectrometry in the environmental sciences. *Anal. Bioanal. Chem.* 403, 1251–64. doi:10.1007/s00216-012-5844-7

Herrera-Lopez, S., Hernando, M.D., García-Calvo, E., Fernández-Alba, a. R., Ulaszewska, M.M., 2014. Simultaneous screening of targeted and non-targeted contaminants using an LC-QTOF-MS system and automated MS/MS library searching. *J. Mass Spectrom.* 49, 878–893. doi:10.1002/jms.3428

Hogenboom, A.C., van Leerdam, J.A., de Voogt, P., 2009. Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography–hybrid linear ion trap Orbitrap mass spectrometry. *J. Chromatogr. A* 1216, 510–519. doi:10.1016/j.chroma.2008.08.053

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T., 2010. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* 45, 703–14. doi:10.1002/jms.1777

Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* 184, 25–32. doi:10.1016/j.envpol.2013.07.048

Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., Hollender, J., 2009. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 43, 7039–7046. doi:10.1021/es901979h

Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 8, 105. doi:10.1186/1471-2105-8-105

Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–951. doi:10.1007/s00216-010-3608-9

Leendert, V., Van Langenhove, H., Demeestere, K., 2015. Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments. *TrAC Trends Anal. Chem.* 67, 192–208. doi:10.1016/j.trac.2015.01.010

Letzel, T., Bayer, A., Schulz, W., Heermann, A., Lucke, T., Greco, G., Grosse, S., Schüssler, W., Sengl, M., Letzel, M., 2015. LC – MS screening techniques for wastewater analysis and analytical data handling strategies : Sartans and their transformation products as an example. *Chemosphere* 137, 198–206. doi:10.1016/j.chemosphere.2015.06.083

Little, J.L., Williams, A.J., Pshenichnov, A., Tkachenko, V., 2012. Identification of “Known Unknowns” Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectrom.* 23, 179–185. doi:10.1007/s13361-011-0265-y

- Luo, Y., Guo, W., Ngo, H.H., Nghiem, L.D., Hai, F.I., Zhang, J., Liang, S., Wang, X.C., 2014. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci. Total Environ.* 473-474, 619–41. doi:10.1016/j.scitotenv.2013.12.065
- Makarov, A., Scigelova, M., 2010. Coupling liquid chromatography to Orbitrap mass spectrometry. *J. Chromatogr. A* 1217, 3938–45. doi:10.1016/j.chroma.2010.02.022
- Miller, T.H., Musenga, A., Cowan, D.A., Barron, L.P., 2013. Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks. *Anal. Chem.* 85, 10330–7. doi:10.1021/ac4024878
- Miners, J.O., Birkett, D.J., 1996. The use of caffeine as a metabolic probe for human drug metabolizing enzymes. *Gen. Pharmacol.* 27, 245–249. doi:10.1016/0306-3623(95)02014-4
- Munro, K., Miller, T.H., Martins, C.P.B., Edge, A.M., Cowan, D. a., Barron, L.P., 2015. Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid chromatography-high resolution mass spectrometry data. *J. Chromatogr. A* 1396, 33–44. doi:10.1016/j.chroma.2015.03.063
- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* acs.analchem.5b02905. doi:10.1021/acs.analchem.5b02905
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014a. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* doi:10.1021/es5002105
- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014b. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol.* 48, 1811–8. doi:10.1021/es4044374
- van der Aa, M., Bijlsma, L., Emke, E., Dijkman, E., van Nuijs, A.L.N., van de Ven, B., Hernández, F., Versteegh, A., de Voogt, P., 2013. Risk assessment for drugs of abuse in the Dutch watercycle. *Water Res.* 47, 1848–57. doi:10.1016/j.watres.2013.01.013
- Zedda, M., Zwiener, C., 2012. Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools. - *Anal. Bioanal. Chem.* 2493–2502. doi:10.1007/s00216-012-5893-y

Supporting Information

Table S3, “All compounds and respective m/z used for the suspect analysis” is too large for this thesis, however all other supporting tables are included here.

Table S1: All target compounds used for final MS/MS analysis

Compound	m/z
2-hydroxy-carbamazepine	271.1080
4-acetylamino antipyrine	246.1234
4-formylamino-antipyrine	232.1086
Acetaminophen	152.0706
Adenosine	268.1035
Benzoylcegonine	290.1385
Caffeine	195.0876
Carbamazepine	237.1022
Carboxylosartan	437.1480
Cocaine	304.1543
Cotinine	176.0939
Desvenlafaxine	264.1958
Gemfibrozil	250.1564
Irbesartan	428.2320
Ketoprofen	255.1014
Lidocaine	235.1807
Losartan	423.1695
Metoprolol	268.1908
Naproxen	231.1016
Paraxanthine/Theobromine/Theophylline	181.0721
Phenacetin	180.1030
Phenazone	189.1022
Phenytoin	252.0888
Sulfamethoxazole	254.0594
Trimethoprim	291.1454
Valsartan	436.2341

Table S2: *Non-target workflow for MPeaks, impact of each algorithm to the number of peaks remaining in each sample*

Samples		Default parameters, less procedural blank peaks, charge state 1, across triplicate	Keep largest ¹³ C peak	Cluster analysis	Keep largest peak from cluster
IWW	1	502	500	459	292
	2	419	417	389	243
	3	415	415	389	238
	4	457	457	420	262
	5	418	417	385	236
	6	379	374	343	209
	7	482	481	445	267
EWW	1	277	276	250	182
	2	303	303	263	184
	3	415	415	360	227
	4	407	407	357	224
	5	378	378	323	210
	6	398	397	338	226
	7	660	658	576	339

Table S4: *Compounds detected in IWW and EWW using ChemSpider-lookup feature of Sieve*

Compounds detected in IWW	Compounds detected in EWW
2-hydroxy-carbamazepine	Adenosine
Benzoylcegonine	Albendazole
Caffeine	Benzoylcegonine
Carbamazepine	Carboxylosartan
Carboxylosartan	Losartan
Cocaine	Paracetamol
Desvenlafaxine	Phenazone
Ibuprofen	Valsartan
Lidocaine	
Losartan	
Metoprolol	
Naproxen	
Paracetamol	
Theophylline/Paraxanthine/ Theobromine	
Trimethoprim	
Valsartan	

2.3 Discussion of Results

All of the articles in this section exemplify the utility of suspect and/or non-target screening. As explained previously, suspect screening incorporates a number of compounds far larger than those available in target screenings. The big advantage of this strategy is that a much wider variety of compounds can be included and there is no need to initially have the reference standards available. **Scientific Article 2** included a database of more than 1000 parent compounds and metabolites, while **Scientific Articles 3** and **4** used a database of 107 and 250 parent compounds and metabolites, respectively. These databases contained compounds of diverse chemical families: drugs of abuse, antibiotics and both human and veterinary pharmaceuticals. While reference standards are commonly purchased to supplement and confirm tentatively identified compounds, the reliability of the tentative identification is high.

The idea of suspect screening is simple: incorporate a database containing a huge number of compounds (as molecular formulae or exact mass and, when available, fragment ions) and use powerful software to match “hits” between the database and injected samples. However, this simple approach has vast consequences for environmental analytical chemists, especially, as it enables compounds for which reference standards are unavailable to be tentatively identified as well as low abundant species which may have been overlooked by a manual inspection. In **Scientific Articles 2** and **3**, a QTOF instrument was used in MS^E mode which allowed the simultaneous acquisition of high and low collision energy accurate-mass data. These simultaneous acquisitions can be visualised in ChromaLynx (within MassLynx), with *nw*-XICs at the *m/z* of all compounds in the database being performed automatically by the software (at low energy), and fragment ions at high energy. The agreement between the software and the database (and sufficiently low mass error, set by the user) leads to the tentative identification of compounds.

Scientific Article 2 focussed on the investigation of surface and wastewaters from Bogotá, Colombia. Samples were screened for pharmaceuticals belonging to different therapeutic groups, veterinary drugs, X-ray agents, PCPs (preservatives and UV filters), sweeteners, illicit drugs, as well as several metabolites. The compounds most frequently detected were the analgesics/anti-inflammatory drugs acetaminophen, diclofenac, ibuprofen, and lidocaine. The antibiotics clarithromycin and lincomycin, the angiotensin II antagonists valsartan and losartan, the anti-epileptic carbamazepine, the beta-blocker metoprolol, the lipid regulator gemfibrozil, and the X-ray agent iopromide were also found. It is also interesting to note that the metabolites of carbamazepine 10,11-dihydro-10,11-dihydroxy and

benzoylecgonine (the main metabolite of cocaine) were also found. One fascinating aspect of this work was the implementation of a common fragmentation workflow for the identification of several metabolites of losartan (**Figure 4** in **Scientific Article 2**) based on three chromatographic peaks sharing three m/z in their respective XICs: 207, 235 and 190. Closer inspection of these peaks led to the identities of two hydroxyl losartan metabolites as well as losartan carboxylic acid being elucidated.

Scientific Article 3 dealt with the analysis of environmental and wastewaters from two countries (Spain and Italy) using two HRMS instruments (QTOF and LTQ-Orbitrap). A small database, similar to that used in **Scientific Article 2**, containing 107 pharmaceuticals and illicit drugs and metabolites was used in both instruments, to allow a wider scope for the screening, based on the different acquisition modes of the instruments. In general, more compounds could be detected (protonated molecule ($[M+H]^+$) at accurate mass and retention time data when standard available) and identified (at least one fragment ion and $[M+H]^+$ and retention time data when standard available) in the Spanish than Italian samples, with cotinine, gabapentin, oxazepam, temazepam and a metabolite of metamizole (4-aminoantipyrine), not identified in the Italian samples (**Table 2, Scientific Article 3**). Irbesartan (antihypertensive), valsartan (angiotensin II antagonist) and two antibiotics ofloxacin and sulfamethoxazole were the most commonly identified compounds in the wastewater samples of both Spain and Italy, while the illicit drug cocaine and its main metabolite benzoylecgonine as well as the addictive opioid painkillers morphine and codeine were also identified. Caffeine, irbesartan, valsartan and venlafaxine (antidepressant) were also found in a high enough concentration in surface water to be identified by at least one of the two instruments. When looking at the results of the two instruments, there was little difference between those compounds found, but there was a discrepancy in the number of compounds identified, which shows the complementary use of two HRMS instruments. As shown in **Figure 2** of **Scientific Article 3**, the antihyperlipidemic bezafibrate could not be identified (i.e. only the $[M+H]^+$ at a comparable retention time to the standard was found) in an EWW sample using the QTOF, due to its low concentration, while in the same sample after analysis with the Orbitrap, three fragment ions were also found, leading to its identification.

In **Scientific Articles 3** and **4**, an LTQ-Orbitrap was used for suspect and non-target screening. With this instrument, simultaneous acquisition of high and low energy data was impossible, due to the longer cycle time of the Orbitrap compared to the QTOF. Instead, the mass spectrometer operated under data-dependent-acquisition (DDA) mode during the complete chromatographic run, in which both MS and MS² spectra were acquired simultaneously. The instrument was initially set to operate in full-scan ('survey') mode with accurate mass measurements. When an ion exceeded a preset threshold,

the instrument switched to product-ion scan mode in the ion trap part. However, this only meant that compounds of a certain (high) intensity would have product ions recorded, and only the protonated molecule would be seen for lower abundant compounds. Nevertheless, the exact mass capabilities of this instrument did lead to the tentative identification of some compounds (as a protonated molecule), which later could be examined in MS/MS mode for fragment ions. The MS/MS acquisition of this instrument was carried out in the ion-trap, and therefore only nominal mass measurements could be recorded for the product ions, which can lead to some erroneous tentative identifications (furthered below). Even so, this instrument does hold distinct advantages, relating to its high resolving power (up to 100,000 FWHM) and MSⁿ capability, which lends itself to non-target analysis.

It must be noted that while tentative identifications do result in fairly high confidence in identity, without a reference standard, it is impossible to have total confidence. This is touched upon in **Scientific Article 4**, for which only tentative identification was possible as no reference standards were utilised in the study. In this article, the steps applied for removing possible false positives included mass error evaluation, isotopic pattern fit, retention time prediction before final fragment ion analysis after MS/MS fragmentation. In total, 22 parent compounds and metabolites were tentatively identified, primarily pharmaceuticals and antibiotics, but more interestingly, also phenacetin, a withdrawn painkiller - now seen as a cocaine adulterant, three alkaloid metabolites of caffeine (theobromine, theophylline and paraxanthine), the purine nucleoside adenosine and cocaine and its main metabolite benzoylecgonine (**Table 3, Scientific Article 4**). Although some uncommon and unexpected compounds were found using the non-target approach, the false positive removal steps explained above give high confidence to their eventual tentative identification.

One interesting aspect that was not covered in **Scientific Article 4** was the implications of the exclusion of benzocaine from further analysis after being outside the predicted retention time following the suspect screening. For the article, the ANN retention time prediction approach was used (**Scientific Article 6**), with a predicted retention time of 15.83 mins. A second retention time prediction analysis was performed using a log_{K_{ow}} only approach (as in **Scientific Article 5**), which predicted benzocaine would have a retention time of 13.35 mins. Using a retention time window of ±11% (±5.5mins) neither of these values are close to the experimental value of 4.81 mins.

There was nevertheless some doubt over the removal of benzocaine, an anaesthetic also used as a cocaine adulterant and as retention time prediction is not perfect, benzocaine was included in the target list for MS/MS analysis, to ensure that it was a false positive. As seen in **Figure 2.1**, there was

only one main chromatographic peak corresponding to the protonated mass of benzocaine ($166.0863 \pm 5\text{ppm}$) at 4.81. In the respective product ion spectra, there was one major fragment (m/z 120.1). Benzocaine was checked in MassBank, for a study using a 30V collision energy (closest to the 35V of this study) to see what the expected product ions are, and found to be 138.3, 120.1 (major), 94.1 and 77.0. Although one fragment ion is in the MS/MS spectrum, there is not enough confidence to assign benzocaine as tentatively identified.

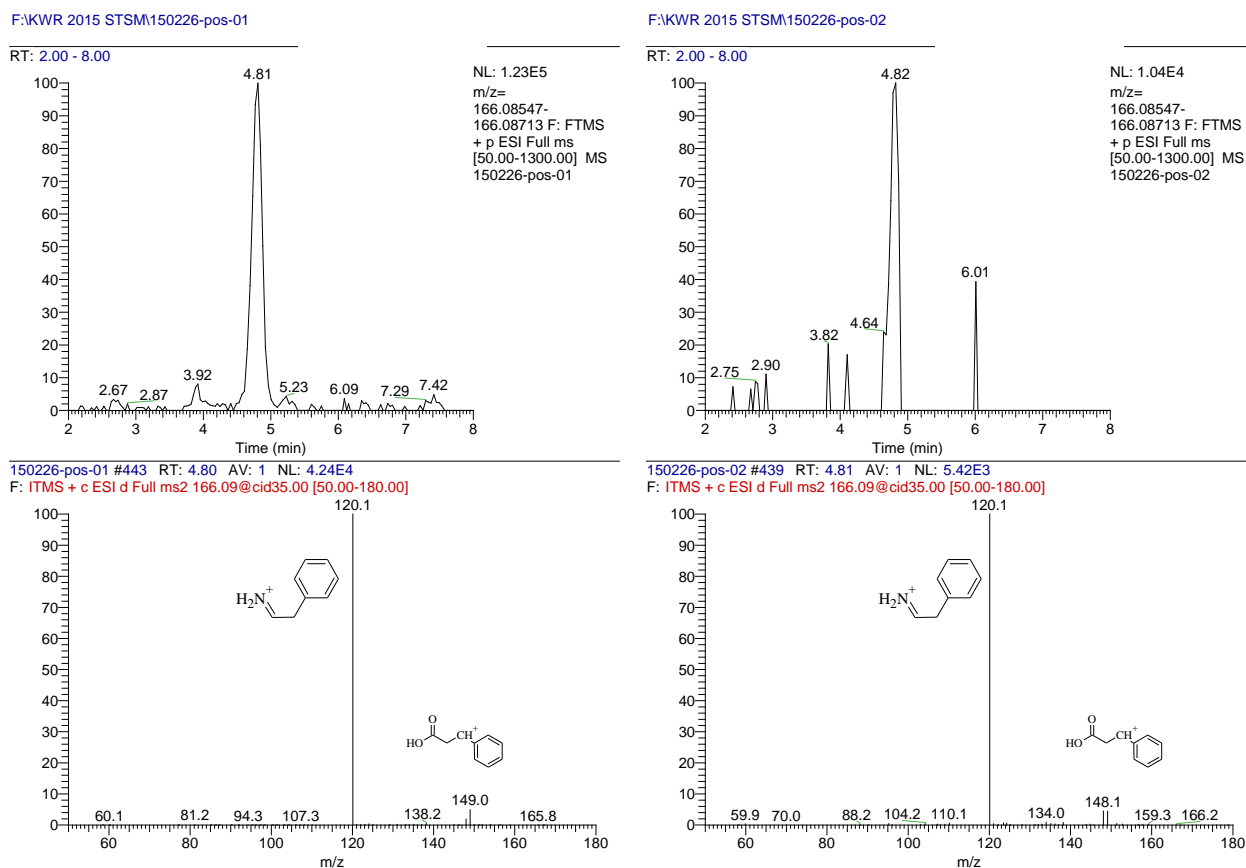


Figure 2.1: Extracted ion chromatogram of $m/z \approx 166.0863 \pm 5\text{ppm}$ with respective mass spectra for the peaks at 4.81 min in EWW (left) and IWW (right). The fragment ions associated with the tentatively identified phenylalanine are shown in the mass spectra.

As benzocaine was found following hidden target screening, the exclusion of it based on inconsistent fragment ions (combined with the predicted retention time data) is enough to totally disregard it for further investigation and accept it as a false positive. It was also not found in the subsequent Sieve hidden target analysis with the in-house database, possibly due to the relatively high intensity threshold used (250,000 for EWW and 62,500 for IWW) and the corresponding intensities of the peak of $m/z \approx 166.0863$ seen in **Figure 2.1**. Nevertheless, the ChemSpider strategy was followed to determine what the peak may have come from. With benzocaine excluded as the source of the peak of $m/z \approx 166.0863$,

a search was made in the ChemSpider database, based on the molecular formula of $C_9H_{11}NO_2$. This formula led to more than 1000 hits, and based on the number of sources, the best candidates were benzocaine and phenylalanine, an amino acid that is used in the biosynthesis of proteins. It is essential in humans, but is not synthesised by the body and must be obtained from the diet, so could plausibly be found in wastewater. Using the ANN model, a predicted retention time of 7.30 mins was found, within the set $\pm 11\%$ window. A search for this compound on MassBank showed that it has one major fragment ion at 120.1, along with a small fragment ion at 149.1. So, phenylalanine could be tentatively identified, based on these fragment ions. Although both benzocaine and phenylalanine have the same major fragment ion (120.1), but different structures (**Figure 2.2**), it is impossible that the same loss occurred in both compounds to give these ions. Further analysis of the structures led to the findings that in the case of benzocaine, a loss of ethanol occurred, while for phenylalanine, it was a loss of formic acid. This points out a limitation of performing MS/MS within an ion trap, where only nominal mass is given. Although ethanol and formic acid have the same nominal mass (46), their exact masses are 46.0419 and 46.0055, respectively, so if an exact mass fragmentation was made, it would have been much easier to determine which compound gave rise to the chromatographic peak.

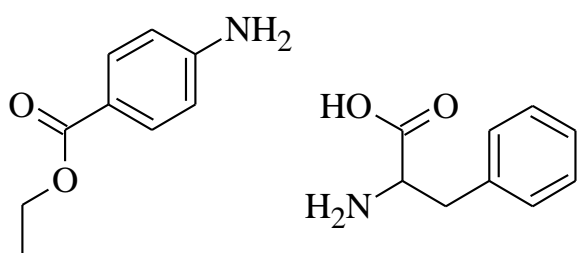


Figure 2.2: Structures of benzocaine (left) and phenylalanine (right)

Phenylalanine does seem a more likely tentative candidate than benzocaine, based on the retention time and fragment ion data. However, a literature search revealed that phenylalanine has never been found in environmental waters, although, being a very common amino acid, its presence should not be discounted. Rather than conform to the “Matthew Effect” phenylalanine can be tentatively assigned as the m/z of 166.0863.

Nevertheless, the conclusions from this investigation are three-fold:

- a) The retention time prediction correctly removed benzocaine as a false positive.
- b) The workflow followed in the article was sound, and correctly led to the removal of benzocaine following the hidden-target analysis. As phenylalanine was not also in the database,

it was not initially considered. However, an additional non-target approach led to its tentative identification.

- c) Without a reference standard, it is impossible to have complete confidence in the identification of a compound. Even with retention time prediction, fragment ions, and a comparative fragment ion spectrum, the lack of studies finding phenylalanine in environmental waters casts doubt over its identification.

One further point regarding the method employed in **Scientific Article 4** is the use of the number of sources to determine which isomer to investigate after the ChemSpider search. This was selected, as explained in the article, as a means to investigate highly consumed/more common drugs and pharmaceuticals, to focus on those of higher commercial importance. This is a known technique to limit the number of possible compounds to focus on, utilised in other non-target methods (Gago-Ferrero et al., 2015; Hug et al., 2014; Little et al., 2012). It does simplify the non-target process, however in doing so, many “potential positive” compounds could be ignored, and, at the same time, some false positives could be included. These false positives are exemplified with the six compounds that made it all the way to the final MS/MS measurement, only to have no comparable fragment ions (2-hydroxy carbamazepine, cotinine, desvenlafaxine, gemfibrozil, irbesartan and phenytoin). Nevertheless, this article did show a way for a non-target (and hidden target) analysis, without the need for any reference standards, and from the steps employed, a high confidence in the tentatively identified compounds was gained.

CHAPTER 3

RETENTION TIME PREDICTION IN THE IDENTIFICATION OF EMERGING CONTAMINANTS IN THE AQUATIC ENVIRONMENT

3.1 Introduction

The detection and identification of emerging contaminants in the aquatic environment is of increasing concern. The current technique of choice for an initial wide-scope screening for qualitative detection and identification is liquid chromatography coupled to high resolution mass spectroscopy (LC-HRMS). As was explained in Chapter 2, there are three screening methods: target, suspect and non-target. In a non-target and suspect screening, either a computationally derived or a manually inputted list of compounds, which contains the exact mass of the compounds of interest, is used as a basis for detection and subsequent identification. Ideally, there would just be one peak arising from the exact mass of the compound of interest in the extracted ion chromatogram (XIC), however, with thousands of organic compounds potentially present in the aqueous environment, there is a high chance that isobaric and isomeric compounds are also present in the same XIC, thereby complicating identification. Several approaches can be used to eliminate unwanted peaks, such as mass window filtering, mass defect analysis and isotopic pattern fit, however, this totally ignores the initial separation by LC, which is also an important parameter (Croley et al., 2012).

Retention time (t_R) prediction does take this important aspect into account, and therefore should be used as an additional means to gain confident identification of compounds. It is best incorporated at the beginning of the identification process, to aid in the removal of false positive peaks and is especially useful in the aforementioned scenario when more than one chromatographic peak is visible at a certain mass. In spite of the improvements in t_R prediction, it is impossible to have an absolutely precise method, so a time window should be incorporated to aid in the identification of possible peaks of interest. The width of this window depends on the initial development and optimization of the prediction method used. For example, a simple approach consisting of only one molecular descriptor would generally have a lower accuracy than a more complex Quantitative structure-retention relationships (QSRRs) approach, necessitating a wider window.

Liquid chromatographic t_R prediction was pioneered by Kaliszan in the late 1970s (Kaliszan, 1977) but only in the past two decades has the evolution and capability of computational techniques enabled serious research. QSRRs represent a powerful tool in chromatography (Kaliszan, 1992), which correlate chromatographic parameters with molecular descriptors, thereby describing the analyte's chemical structure (Giaginis and Tsantili-Kakoulidou, 2012). There are many different QSRR approaches, with linear solvent energy relationships (LSERs) and, lately, support vector machines (SVM), random forests (RF) and artificial neural networks (ANNs) the most popular.

The current work on LSERs is based on the research of Abraham *et al.*, with the following equation representing LSERs for LC, which describes the phase transfer of a substance between two different phases (Abraham *et al.*, 2004):

$$\text{Solute property} = aA + bB + sS + eE + vV + c$$

where solute property can be of any kind, such as $\log K_{OW}$ and $\log P$. A, B, S, E and V are compound descriptors, while a, b, s, e and v are phase descriptors describing the matrix/solvent interactions and c is the intercept of the equation (Ulrich *et al.*, 2011). The terms aA and bB represent the hydrogen bond interactions, representing the acidity and basicity respectively. The term sS reflects the dipolarity/polarizability, eE is the excess molar refraction and vV is the McGowan volume. The outcome of a LSER analysis is a set of regression coefficients which provide information about which solute–solvent interactions significantly affect the retention process. It should be mentioned that the regression coefficients are interrelated and do not carry independent information. This model is known to contribute primarily to the general understanding of a partition process and less to optimise separations, but is usually not accurate enough for prediction purposes (Héberger, 2007). Another major problem is that all values are needed and any missing value can be problematic (Livingstone, 2003).

SVM, RFs and ANNs have all been utilised and compared in numerous studies (Fatemi *et al.*, 2010; Goudarzi and Shahsavani, 2012; Hancock *et al.*, 2005; Héberger, 2007; Ibáñez *et al.*, 2016; Shinoda *et al.*, 2008), but no one can be considered distinctly superior.

SVM are a group of supervised learning methods that can be applied to classification or regression and have gained popularity among computational chemists, particularly for drug design. The QSRR application of SVM regression, however, are rare, despite its potential as a viable alternative (Ivanciuc, 2007). The main advantage seen for SVM (particularly over ANNs) is that the strong theoretical background provides a high generalization capability and can avoid local minima, i.e. it has the ability to accurately predict for new data (Shinoda *et al.*, 2008). On the other hand, while its predictive capacity is great, for simple datasets, the solution can be overfitted (Ivanciuc, 2007).

RF is a set of individual classification and regression trees which are combined to provide a prediction tool with high performance, introduced by Breiman (Breiman, 2001). In the RF algorithm, each tree is grown on a separate training set that is a bootstrap replicate of the original data. In each bootstrap sample, with the same length as the original data, some of the original data is repeated, and the left out samples are called “out-of-bag”. To construct a tree, at each node, a small random subset of p

descriptors or explanatory variables ($m < p$) is chosen, and the best split of the feature space is done using only this subset (Goudarzi and Shahsavani, 2012). Each tree is then grown to the maximum size and is not pruned. It has been shown that as the number of trees increases, the test error always converges. By using many trees, RF have the ability to identify important variables and their relationship with the response. The number of trees, therefore, needs to be set sufficiently high to allow for this convergence as well as to identify the most important variables (Goudarzi and Shahsavani, 2012; Hancock et al., 2005). This randomness has the effect of building new trees with different structures, increasing the variety of relationships modelled within the forest, which in turn improves the overall predictive performance and makes it robust against overfitting. However, while RF are good for identifying main trends, it can lead to some points being poorly predicted (Hancock et al., 2005).

ANNs mimic the neurons of the human brain's way of working in acquiring knowledge from its environment, by "learning" from a set of case examples in the training phase (Giaginis and Tsantili-Kakoulidou, 2012; Kaliszan et al., 2003). There are three phases in the optimisation of ANNs: training, validation and blind. In the first, training phase, a list of appropriate molecular descriptors as well as the true output value is needed (i.e. retention time that will be later predicted). Simultaneously, a second dataset of case examples is used for verification and to evaluate the overall ANN predictive error. In this set, the true output values are not used for learning and the number of training cycles can be stopped by the user or when the software determines the measured error is at its minimum. In many cases, this is where the optimisation and development of the ANN would end, however, sometimes a third, blind test, dataset can be used, where the true output value is not needed, to test the predictive power of the network. The main disadvantage seen for ANN is a tendency for overfitting (i.e. giving a misleadingly accurate set of predictions) and the need to select variables in advance. (Goudarzi and Shahsavani, 2012; Héberger, 2007)

While QSRRs sound quite straightforward in theory, in practice there are some key points to consider; with one crucial aspect being which molecular descriptors should be used and how they should be calculated (Héberger, 2007). While some techniques (SVM and RF) are able to incorporate variable selection as well as predictive analysis, ANN and simple linear regression require prior selection of variables. Simplified molecular-input line-entry system (SMILES) was developed in the late 1980s as a way in which to easily transfer and encode molecular structures using line notation (Weininger, 1988). Together with mol files (Dalby et al., 1992) they have become one of the quintessential means

for gaining molecular descriptors of compounds from software such as Dragon (Talete S.R.L., 2013) and VCCCLabs (Tetko et al., 2005).

There are thousands of potential molecular descriptors, including physicochemical, topological, electro-topological and energy parameters that could be used for QSRR calculations, leading to difficulties in deciding which to use. Physicochemical descriptors are generally strongly correlated to retention, and are therefore more widely used as a starting point for retention time prediction, however they are often not experimentally available and predicted values differ widely (Livingstone, 2003). To eliminate *a priori* knowledge from impacting on the prediction capabilities, and to derive the most relevant molecular descriptors, various methods can be used, such as stepwise correlation and genetic algorithms (Pourbasheer et al., 2015; Shen et al., 2003). Barron *et al* have also looked into the impact of various descriptors in the calculation of a retention time prediction model, with logP and pKa expectedly the most important (Barron et al., 2009; Miller et al., 2013). It is imperative to note that the descriptors found to be of the greatest influence will differ depending on the LC system, gradient, additives and column. Therefore, all descriptors will have to be rechecked if any aspect of the LC methodology changes.

In order to avoid these issues, and to have a simpler, albeit less precise, t_R predictor, a single descriptor can be used together with retention time, such as $\log K_{ow}$, which is known to represent a reasonably good model for retention of neutral compounds on C_{18} columns with methanol as the organic modifier (Kern et al., 2009). K_{ow} itself is a complex property, which expresses the hydrophobicity of neutral organic chemicals as their octan-1-ol/water partition coefficient. Ideally, for a reversed-phase liquid chromatography system the capacity ratio ($\log k'$) of a compound obtained with a mobile phase consisting of pure water would best represent its hydrophobicity (Braumann, 1986). It has been shown that the relationship between $\log k'$ and the volume fraction (φ) of the organic fraction gives the following equation, which can then be used as an indicator of K_{ow} (De Voogt et al., 2012):

$$\log k' = \log k_0 - b\varphi$$

The slope $-b$ represents the decrease in retention of the solute as a function of increasing organic modifier content. This slope depends on the chromatographic conditions (type of column, chemical composition, temperature, back pressure, etc.), and it can be postulated that a $\log k'$ value obtained on a C_{18} column with a methanol-water mobile phase could be used at single value of φ (de Voogt et al., 2012). The robustness of the relationship between C_{18} retention time and $\log K_{ow}$ is shown in the

use of C₁₈ columns in the Organisation for Economic Co-operation and Development (OECD) Guidelines for the Testing of Chemicals for the estimation of LogK_{ow}(OECD, 2004).

This simplistic approach linearly correlates the retention time and logK_{ow} of the studied compounds, giving an equation that can be used to predict further compounds. The big advantage of using this approach is its simplicity to implement and that fewer compounds are needed for prediction. With the less precision associated with this method of retention time prediction, it is generally applied as an additional rather than stand-alone means of detection/identification. Nevertheless, its use as a means of retention time prediction is not necessarily worse than incorporating multiple descriptors.

While research on *t*_R prediction and the detection/identification of emerging contaminants in the aquatic environment has been made separately, the unification of these two topics is of great current and future interest. Currently, *t*_R prediction is based on a laboratory's own chromatographic system and is therefore not easily transferable. However, recently, Barron and McEneff have shown that ANNs can be used to successfully predict the retention times of more than 1000 structurally and chemically diverse molecules in a range of complex matrices (Barron and McEneff, 2016), while PredRet has been introduced as a community-sharing approach for *t*_R prediction, which uses a direct mapping approach for the prediction of *t*_R across different chromatographic systems (Stanstrup et al., 2015). This shows that the future is looking bright for this area and with computational and instrumental advances, this research will continue to flourish.

3.2 Scientific Articles

This chapter includes two articles, “*Critical evaluation of a simple retention time predictor based on LogK_{ow} as a complementary tool in the identification of emerging contaminants in water*” (**Scientific Article 5**) and “*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*” (**Scientific Article 6**). The first concerns the development and evaluation of a very simple retention time prediction method, based solely on logK_{ow} and retention time, while the latter incorporated a more complex artificial neural network (ANN) approach.

Scientific Article 5 was envisaged as an example of how a research group could initially start investigating and using retention time prediction. It was not proposed to be a perfect predictor, rather a simple approach that any and every laboratory could use. All that is needed for this approach is reference standards and a reliable LC instrument. Nearly 600 standards, including pesticides, human and veterinary pharmaceuticals, drugs of abuse and mycotoxins were used for this work, however, similar works have incorporated as little as 88 (Nurmi et al., 2012) and 92 (Kern et al., 2009) standards, with similar results. These standards were injected in an UHPLC-QTOF instrument and the logK_{ow} values were searched on the freeware ALOGPS 2.1 (Tetko et al., 2005). Three subsets were made to test the predictor: all compounds (n=595), pesticides (n=350) and non-pesticides (n=245), with the first two showing the greatest promise and were later used to test real samples. A comparison was also made with mass window to see whether reducing the mass window also reduces the number of peaks in the extracted ion chromatogram.

Scientific Article 6 used a slightly smaller subset (544) of the standards of **Scientific Article 5** and an alternative, QSRR method; employing artificial neural networks (ANNs). To test network performance, training (n=344), verification (n=100) and blind (n=100) sets were used. For all datasets (n=544), 85% of all compounds could be predicted within two minutes of the measured retention time, while this increased to 90% for the verification set (n=344) and 95% for the blind set (n=100). The accuracy for the blind set was very good, leading to further blind testing with metabolites and compounds for which no reference standards were available, which led to the tentative identification of eleven compounds. As with **Scientific Article 5**, an investigation was made into the number of peaks able to be disregarded after incorporating the retention time window.

The results of the two studies in this section have been published:

Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water R. Bade, L. Bijlsma, J. V. Sancho, F. Hernández, *Talanta* 139, **2015**, 143-149.

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 R. Bade, L. Bijlsma, T. H. Miller, L. P. Barron, J. V. Sancho, F. Hernández, *Science of the Total Environment* 538, **2015**, 934-941.

3.2.1 Scientific Article 5

**CRITICAL EVALUATION OF A SIMPLE RETENTION
TIME PREDICTOR BASED ON LOG_{KOW} AS A
COMPLEMENTARY TOOL IN THE IDENTIFICATION
OF EMERGING CONTAMINANTS IN WATER**

Richard Bade, Lubertus Bijlsma, Juan V. Sancho, Felix Hernández

Talanta (2015) 139: 143-149



Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water



Richard Bade, Lubertus Bijlsma, Juan V. Sancho, Felix Hernández*

Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain

Abstract

There has been great interest in environmental analytical chemistry in developing screening methods based on liquid chromatography-high resolution mass spectrometry (LC-HRMS) for emerging contaminants. Using HRMS, compound identification relies on the high mass resolving power and mass accuracy attainable by these analyzers. When dealing with wide-scope screening, retention time prediction can be a complementary tool for the identification of compounds, and can also reduce tedious data processing when several peaks appear in the extracted ion chromatograms. There are many *in silico*, Quantitative Structure-Retention Relationship methods available for the prediction of retention time for LC. However, most of these methods use commercial software to predict retention time based on various molecular descriptors. This paper explores the applicability and makes a critical discussion on a far simpler and cheaper approach to predict retention times by using LogKow. The predictor was based on a database of 595 compounds, their respective LogKow values and a chromatographic run time of 18 minutes. Approximately 95% of the compounds were found within 4.0 minutes of their actual retention times, and 70% within 2.0 minutes. A predictor based purely on pesticides was also made, enabling 80% of these compounds to be found within 2.0 minutes of their actual retention times. To demonstrate the utility of the predictors, they were successfully used as an additional tool in the identification of 30 commonly found emerging contaminants in water. Furthermore, a comparison was made by using different mass extraction windows to minimize the number of false positives obtained.

Keywords: Retention Time Prediction, Liquid Chromatography, Time-of-flight mass spectrometry, Water, Emerging Contaminants, Pesticides

1. Introduction

Many environmental chemists have focused their research on “emerging contaminants” in water, which encompass a wide-range of compounds including drugs for health and personal care, flame retardants, illicit drugs and all kinds of transformation or by-products [1]. These compounds can have some detrimental effects on the environment [2] and it is therefore important to perform frequent monitoring on these compounds in order to know their concentrations and fate in water.

High resolution mass spectrometry (HRMS), using analyzers such as Quadrupole Time-of-Flight (QTOF)-MS and Linear Ion Trap (LITQ) Orbitrap MS, has revolutionized screening of emerging contaminants [3], [4], [5] and [6]. It offers the possibility to investigate the presence of theoretically unlimited compounds once the analysis has been performed and data acquired, and considering their compatibility with the requirements of the chromatographic separation and MS ionization. Due to their high resolution, both hybrid instruments provide data with high mass accuracy and frequently allow tentative identification of compounds even without having reference standards [7] and [8].

Identification of the compounds detected is obviously facilitated when reference standards are available, as relevant information on retention time and fragment ions is included. However, it is also possible to perform screening without the need for reference standards, simply on the basis of a large database, where empirical formulae (i.e. exact mass) are the only information required. Post-target screening without standards (i.e. suspect) is becoming more and more common. Here, the exact mass of the compounds of interest are gathered, then searched and extracted from HRMS spectra, using a narrow mass window (20 mDa), in order to find potential positives [5], [7], [9], [10] and [11]. The increased resolving power of modern TOF and Orbitrap allows an even narrower mass window, reducing matrix interferences, leading to a cleaner chromatogram. Ideally, there would be a single peak in the chromatogram, coming from the suspect. However, in more complex environmental matrices, it is likely that there is more than one peak in the eXtracted Ion Chromatogram (XIC), arising from various isobaric or isomeric compounds, complicating the identification of the compound of interest. A variety of techniques can be used to aid in the detection/identification process, such as mass window filtering, mass defect analysis or isotopic pattern fit.

With the high quality data obtained by HRMS, compound separation by liquid chromatography (LC) is sometimes overlooked, while it is also an important parameter [12]. Retention time prediction does take this LC aspect into account in the identification process and can therefore be a useful technique when performing a wide-scope screening. Reliable information on retention time can focus the identification process solely on those peaks that are in agreement with the predicted retention time, ignoring other false positives that may appear in complex matrix samples which do not correspond to the candidate under research.

There are several *in silico*, Quantitative Structure–Retention Relationship (QSRR) methods available for the prediction of LC retention times, used in a variety of research [13] and [14]. The principal aim of QSRR is to predict retention data from the molecular structure, using descriptors such as molecular mass, polar surface area, $\log P$, molar polarisability and molar volume. Linear Solvation Energy Relationship (LSER) has been also proposed for retention time prediction. LSERs analyze any free energy related property by five fundamental solute parameters: the hydrogen bond acidity and hydrogen bond basicity, excess molar refraction, dipolarity/polarizability and the logarithm of the gas–hexadecane partition coefficient or the characteristic McGowan volume [14]. The major problem for LSERs is that all terms are needed and any missing value can be problematic [15]. While work with LSERs in this area is ongoing, Artificial Neural Networks (ANNs), a predictive computing technique, has shown itself as a promising retention time predictor [16].

In the presented work, a free and much simpler approach was tested using only LogKow for retention time prediction. By utilizing this predictor, we show a way to reduce the amount of time spent poring through chromatograms, with many peaks in the chromatogram able to be disregarded. A critical discussion is made on this predictor, showing the advantages and drawbacks. The usefulness of mass window filtering as a complementary tool to eliminate peaks not arising from the compound of interest is also evaluated in order to facilitate the identification process of emerging contaminants in water in wide-scope screening procedures.

2. Experimental

2.1. Reagents and chemicals

A total of 595 reference standards were used in the development of the retention time predictor. This combined retention times of 311 and 284 individual standards and their retention times from an in-house database and a Waters database (Waters, Milford, MA, USA), respectively. See Table S1 for a list of all compounds used, retention time and LogKow. There were some duplicates between the two sets of standards, which were used to ensure the consistency in retention time (see Section 3.1 for more information).

Details relating to the standards can be found elsewhere [17]. Retention times were obtained by injecting mixed working standard solutions (25 $\mu\text{g/L}$ or 50 $\mu\text{g/L}$, diluted from mixed standard solutions in methanol or acetonitrile with water).

2.2. Water samples

Four influent wastewater samples (24 h composite samples) were collected from different wastewater treatment plants (WWTPs) of Spain and Germany, in March, 2013. In addition, grab samples were taken from 11 surface waters from several points located in Spain and Colombia between November, 2010 and May, 2013. All these samples had previously been used in different studies performed at our lab using UHPLC–QTOF MS for their analysis. Sample treatment was based on solid phase extraction using polymeric Oasis HLB cartridges, which are able to retain organic compounds within a wide range of polarity.

2.3. UHPLC–QTOF MS

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (XEVO G2 QTOF, Waters Micromass, Manchester, UK), using a ESI (Z-Spray) interface operating in positive ion mode. The chromatographic separation was performed using an Acquity UPLC BEH C18 100× 2.1 mm², 1.7 μm particle size column (Waters) at a flow rate of 300 μl/min. The mobile phases used were A = H₂O and B = MeOH, both with 0.01% formic acid. The initial percentage of B was 10%, which was linearly increased to 90% in 14 min, followed by a 2 min isocratic period and, then, returned to initial conditions during 2 min. The total run time was 18 min. Nitrogen was used as drying gas and nebulizing gas. TOF-MS resolution was approximately 20,000 at full width half maximum at m/z 556. MS data were acquired over an m/z range of 50–1000, at 0.4 s scan time. A capillary voltage of 0.7 kV and cone voltage of 20 V were used. Collision gas was argon 99.995% (Praxair, Valencia, Spain). The desolvation temperature was set to 600 °C, and the source temperature to 135 °C. The column temperature was set to 40 °C. MS data was acquired in MS^E mode, selecting a collision energy of 4 eV for low energy (LE) and a ramp of 15–40 eV for high energy (HE) [5] and [18].

2.4. Data processing

MS data processing was performed manually on MassLynx v 4.1 (Waters Corporation), looking at the raw data in chromatogram view, initially with a mass extraction window of 20 mDa using the retention time predictor developed in this study. Later, different mass extraction windows (50 mDa, 10 mDa and 5 mDa) were also evaluated (see Section 3.3). All peaks above an intensity of 3000 were counted.

3. Results and discussion

3.1. Development of retention time predictor

A dataset of 595 compounds was used to initially prepare the retention time predictor. The retention times for the compounds from the “Waters” database were obtained using the same column but a slightly different gradient from the one described in this paper, however the 94 common compounds between the datasets were compared using a linear correlation on Excel ($R^2=0.9279$). These compounds cover the LogKow range of -1 to 8 , therefore comprising the great majority of the LogKow range for the compounds under investigation. The equation from this correlation was used to convert all the Waters retention times to fit our in-house retention times (Fig. S1). The LogKow of each of these compounds was estimated using the freely available ALOGPS 2.1 software (VCCLAB, 2005 [19]). A linear correlation was again made to compare the LogKow and retention time (Fig. 1).

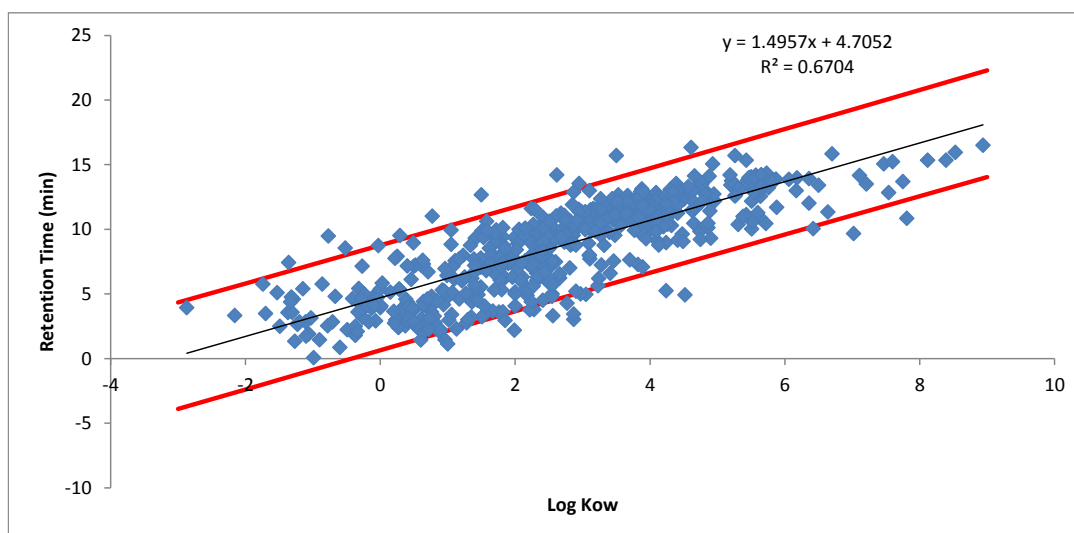


Fig. 1: Correlation between retention time and LogKow for all compounds, with 95% confidence interval

In a similar study, Nurmi et al. [11] investigated the relationship between LogKow and retention time to help reduce the number of candidates in a post-target screening method, using calculated LogKow values. A linear regression for 88 compounds was made, giving an R^2 of 0.63, and a retention time window of 5 min over the chromatographic run of 16 min was used. Likening this work to the current, in spite of having a far larger dataset, the correlation coefficient does not differ markedly. Kern et al. [9] exploited this relationship in the analysis of transformation products (TPs) of organic contaminants. Specific compounds were used for the training set, with only standards that were estimated to be predominately neutral at an elution pH of 3 deemed acceptable, to reduce the number of false negatives for both neutral and ionic TPs as the latter are less retained than their corresponding neutral species. Nevertheless, the correlation for the 92 reference standards was very

good, with an R^2 of 0.87. However, the current work is more wide-reaching, containing nearly 600 compounds of differing classes and physicochemical properties and is therefore expected to have a worse correlation.

In order to improve this correlation, the 595 compounds were subdivided, not based on physicochemical parameters, solely on class, into two smaller subsets: pesticides and non-pesticides. The dataset initially contained pesticides, drugs of abuse, antibiotics, pharmaceuticals, veterinary drugs and mycotoxins. The vast majority of the compounds were pesticides (345 compounds), which made up one subgroup, while all the other compounds made up the other (250 compounds) (Figs. S2 and S3).

The “pesticides only” grouping did make for a scarcely better correlation ($R^2=0.6947$), while the “non-pesticides” had a slightly worse correlation ($R^2=0.6518$), compared with the overall correlation ($R^2=0.6704$). Although these correlations have rather large variability (especially for “non-pesticides”), it was thought to compare the predicted retention time (made from the equations, where “ x ” is the LogKow, in each of the figures) with the actual retention time for each compound. It was expected that there would be some deviation between the experimental and predicted retention times because it is difficult for the algorithm used by the LogKow predictor (ALOGPS 2.1) to cope with complex molecules, leading to some inaccurate LogKow values. Table 1 shows the differences between the predicted and actual retention times for the three cases.

Table 1: Comparisons between predicted and actual retention times

RT window (\pm minutes)	Prediction for all compounds (% of all compounds found within RT window)	Prediction for subsets (% of all compounds found within RT window)	
		Non-Pesticides (n=250)	Pesticides (n=345)
	All (n=595)		
0.5	18	15	30
1	37	34	53
1.5	55	49	68
2	68	65	79
2.5	81	73	86
3	88	83	90
3.5	92	89	93
4	96	90	95
4.5	97	93	96
5	98	95	97

In spite of the variability of the data, the retention times of approximately 95% of the compounds were found within 4 min of their actual retention time. In the case of the subsets, 79% of the pesticides can be found within 2 min of their actual retention time, which is 14% better than for non-pesticides.

From these results, it was thought to use ± 2 min window ($\pm 11\%$ of the chromatographic run) for use in real samples. The “pesticides only” predictor was selected for pesticides and “all compounds” for the other compounds, wherein approximately 80% and 70% of the compounds will be respectively found.

This predictor was designed specifically for this precise chromatographic system, gradient and method. If applied to other separation systems, the correlation between LogKow and retention time for these particular compounds may differ widely. However, it is easy to adapt this retention time predictor to other systems using the methodology outlined in Section 3.1. Some groups have introduced a retention time index [6] to cope with this limitation. In order for a more complete predictor, training/validation sets comprising different compounds could be incorporated. Furthermore, the evaluation of different chromatographic conditions such as different columns and gradients could be carried out.

3.2. Impact of experimental and predicted LogKow

In an attempt to gain narrower retention time windows, a study was made on the impact of experimental versus predicted LogKow values. A study was made comparing the accuracy of predictive software, wherein ALOGPS 2.1 was shown to be quite accurate and only differed from the measured values by up to 0.5 log units [15]. Software such as ALOGPS 2.1 is prone to systematic errors, especially for complex molecular structures, because correction factors for certain structural configurations might be missing [9]. To test the difference between experimental and predicted LogKow for the retention time predictors, “experimental” LogKow values were found for 280 of the pesticides in the Pesticide Manual [20] and for 52 drugs of abuse, antibiotics, veterinary drugs and pharmaceuticals on DrugBank [21].

Fig. 2 shows the difference between the experimental and predicted values for the 332 compounds at intervals of 0.1 log units. As seen in the figure, fewer than 2% (corresponding to 5 of the 332 compounds) had the same experimental and predicted LogKow. Cumulatively, 77% of the compounds had an absolute difference within 1 log unit; however 10% still had a difference greater than 2 log units.

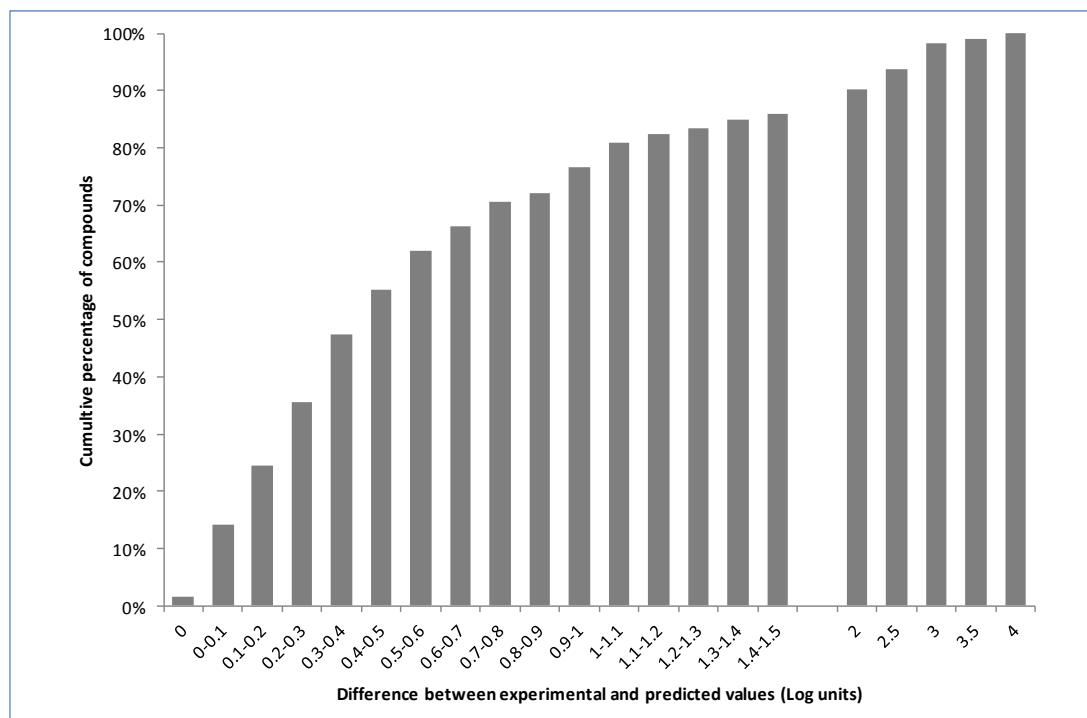


Fig. 2: Difference between experimental and predicted LogKow values and cumulative percentage of compounds within each differential interval.

These findings alone show that while predicted values give a good estimate, the fact that some compounds had a difference more than two Log values shows that experimentally derived values are preferable for any latter retention time predictions. The 36 compounds whose values of experimental and predicted LogKow differed by greater than 2 log units were removed from the overall compound list to see the impact on the correlation coefficient. It was found that the R^2 did not differ, with only a change from 0.6704 to 0.6737 following their removal (Fig. S4).

The difference between experimental and predicted LogKow values had such little impact on the correlation between LogKow and retention time, which gives credence to the use of predicted values. Furthermore, experimental values are not possible for all compounds and predicted values open up the possibility to work with new emerging contaminants and TPs, which is of particular relevance when investigating organic micropollutants in waters.

3.3. Application of the retention time predictor to real water samples

Fifteen influent wastewater and surface water samples were selected, representing different types of water sources, to test the retention time predictors. As stated previously, two predictors were used: one for pesticides only and one for non-pesticides. A set of 30 compounds were selected, based on their prevalence in environmental water samples [22], [23] and [24] and their retention times were predicted using the aforementioned equations with the predicted LogKow of each of the compounds (Table S2).

A retention time window of ± 2 min (from the predicted value) was given for each compound, and they were searched with a mass window of 20 mDa. All peaks in the narrow window (nw)-XIC for each compound were counted manually (above intensity of 3000), while all peaks outside the prediction window were disregarded.

All of the 30 compounds were detected in the water samples. Of these, 20 were found in the ± 2 min retention time window (Fig. 3). Remarkably, eight of the 20 compounds had only one peak inside the retention time window. In addition, the percentage of peaks outside the retention time window of ± 2 min (and therefore not pertaining to the compound of interest) was found to be 35%, meaning that over one-third of the peaks in the XICs could be disregarded through retention time prediction. This retention time prediction shows a noticeable reduction in time spent processing data for potential positive samples.

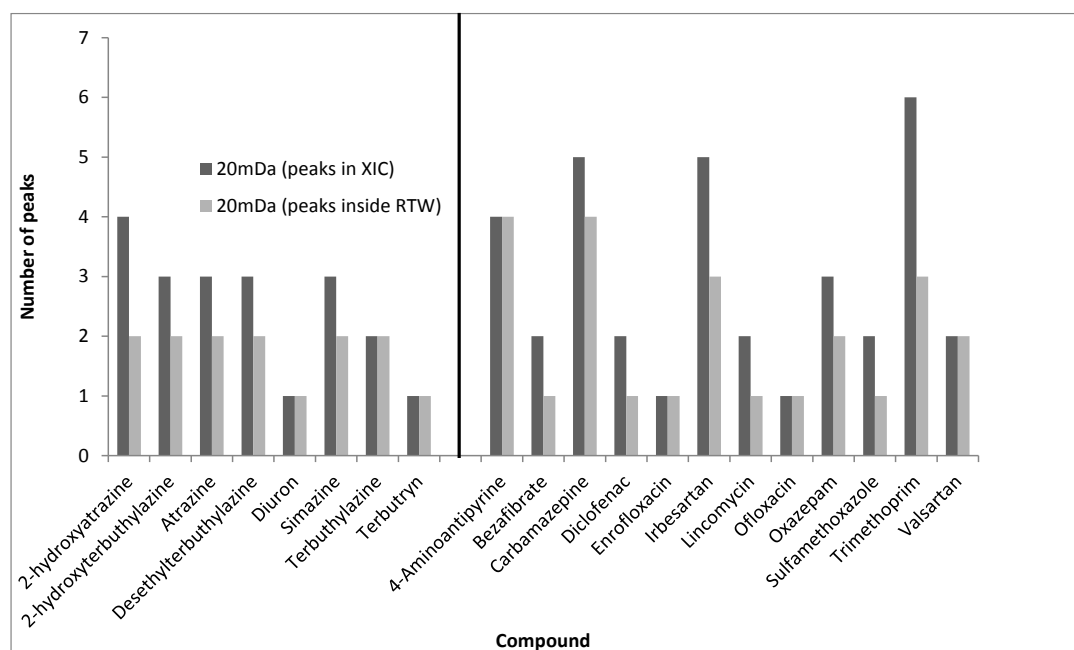


Fig. 3: All 20 compounds found within the ± 2 minute retention time window and their associated peak data. The compounds on the left are all pesticides while those on the right are pharmaceuticals. The dark grey lines represent the average number of peaks found in the XIC of each compound and the light grey lines represent the average number of peaks inside the retention time window.

3.4. Complementary use of mass chromatogram extraction window

Using a mass window of 20 mDa, four compounds only had one peak in the XIC (Fig. 3). To complement the information and applicability of the retention time predictor, a comparison was made with three additional mass chromatogram extraction windows (50 mDa, 10 mDa and 5 mDa). An extra window of 20 mDa was used as a reference for these tests.

The 10 extra compounds, whose predicted retention time fell outside of the ± 2 min window, were also included in this test (see Table S2). The results of this comparison are shown in Table 2. As expected, the number of peaks observable in the XIC decreases as the extraction window decreases, from 119 (50 mDa) down to 62 peaks (5 mDa). However, it is of note that even at the 5 mDa extraction window and within the retention time window, there were still unknown peaks not just pertaining to the compound of interest. In these situations, retention time prediction is very helpful in the reduction of false positive findings.

Fig. 4 shows the influence of mass windows in XICs and the predicted retention time window for two compounds detected in surface water (benzoylecgonine, BE) and influent wastewater (trimethoprim). In the case of BE (major metabolite from cocaine), using a 50 mDa XIC, several isobaric compounds are also seen; however, by decreasing the mass extraction window, all of these peaks disappear, leaving just the peak of BE at 4.53 min and a small spike at 6.5 min. Although BE was able to be identified solely with a nw-XIC, the retention time fell just outside the predicted window (2.48 ± 2 min). As stated above, the use of ± 2 min retention time window led to a 70% success rate in the predictions for the “all compounds” group, where BE is included. Using a slightly wider window such as ± 2.5 min to get a success rate of 80%, similar to the pesticides group (see Table 1) would have led to BE being inside the prediction window. In any case, it seems that improvements are needed in the retention time prediction in order for it to be more useful in the identification process.

Table 2: Number of peaks found in the chromatograms (XICs) of 50mDa, 20mDa, 10mDa and 5mDa and inside the predicted retention time window. The lower compounds are those whose empirical retention times fall outside the ± 2 minute predicted retention time window

Compound	50mDa		20mDa		10mDa		5mDa	
	Peaks	Inside RTW	Peaks	Inside RTW	Peaks	Inside RTW	Peaks	Inside RTW
2-hydroxyatrazine	7	3	4	2	3	2	3	2
2-hydroxyterbuthylazine	5	3	3	2	2	2	1	1
Atrazine	4	2	3	2	2	1	2	1
Desethylterbuthylazine	3	2	3	2	2	1	2	1
Diuron	2	1	1	1	1	1	1	1
Simazine	3	2	3	2	3	2	3	2
Terbuthylazine	2	1	2	2	1	1	1	1
Terbutryn	4	2	1	1	1	1	a	a
4-Aminoantipyrine	4	3	4	4	3	2	3	2
Bezafibrate	2	1	2	1	2	1	2	1
Carbamazepine	4	3	5	4	2	2	2	2
Diclofenac	2	1	2	2	2	1	1	1
Enrofloxacin	1	1	1	1	1	1	1	1
Irbesartan	6	3	5	3	3	2	3	2
Lincomycin	3	2	2	1	2	1	3	2
Ofloxacin	2	1	1	1	1	0	1	1
Oxazepam	4	3	3	2	3	2	2	2
Sulfamethoxazole	3	1	2	1	2	1	1	1
Trimethoprim	7	3	6	3	6	3	4	2
Valsartan	4	2	2	2	1	1	1	1
Total peaks	76	43	57	37	45	29	39	29
Carbendazim	4		3		2		2	
Desethylterbumeton	7		4		3		3	
Imazalil	4		2		2		2	
Codeine	5		5		3		2	
Paracetamol	4		4		4		4	
Temazepam	7		4		3		3	
Venlafaxine	5		5		3		3	
Benzoylcgonine	3		3		2		2	
Cocaine	3		2		2		2	
Methamphetamine	5		3		3		2	
Total peaks	119		90		70		62	

a: not found in this mass extraction window

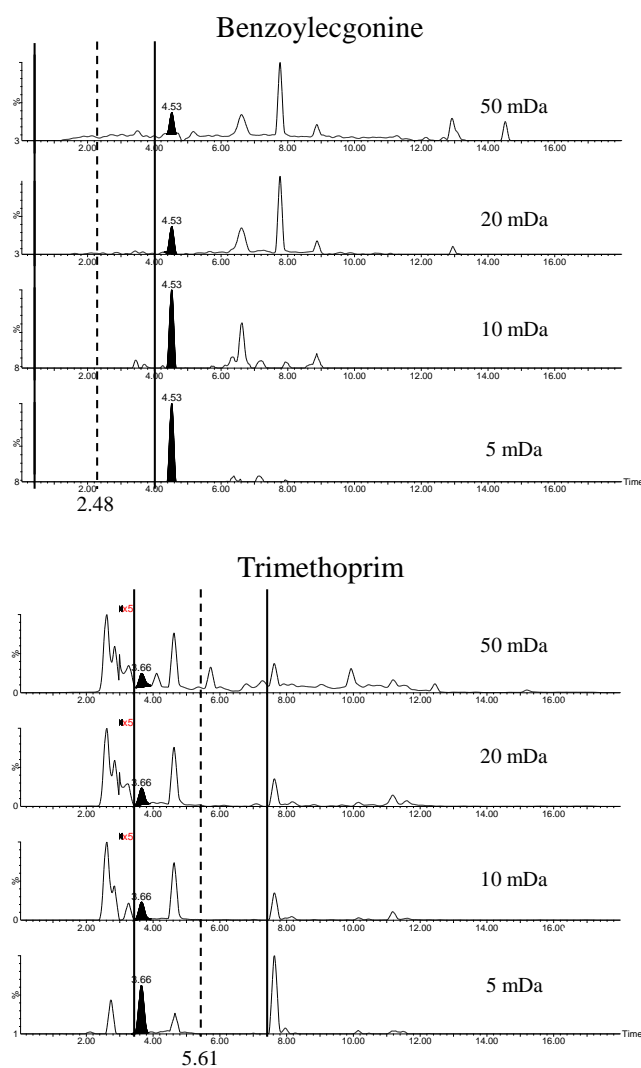


Fig. 4: TOP: XIC of Benzoyllecgonine ($m/z \approx 290.139$, 4.53 min) at 50mDa, 20mDa, 10mDa and 5mDa, with related predicted retention time window (2.48min \pm 2 minutes) in surface water. BOTTOM: Trimethoprim ($m/z \approx 291.146$, 3.64 min) and related retention time prediction window (Predicted retention time 5.61 min \pm 2 minute window) at XICs of 50mDa, 20mDa, 10mDa and 5mD in influent wastewater. XICs at 10mDa, 20mDa and 50mDa all have been 5 times magnified (from 3 – 16 min) to highlight other peaks.

In the case of the antibiotic trimethoprim, the peak is easily seen at 5 mDa (3.66 min); however at a larger mass extraction window, two isobaric compounds are observed at a much higher intensity (2.5–3.0 min). With the retention time window (5.61 \pm 2 min), these peaks, as well as the one at 7.6 min are removed, leaving just the two peaks at 3.66 and 4.6 min. The first one corresponded to trimethoprim, while the second was an unknown. This example shows the true utility of retention time prediction, especially in alliance with HRMS. While mass extraction windows can be narrowed to remove some interfering peaks, even at a 5 mDa mass window, pseudo-isobaric interferences remain. By incorporating retention time prediction, some of these false positive peaks can be removed.

4. Conclusions

A critical evaluation has been made on the applicability of retention time prediction based on LogKow to help in identification of suspect compounds in screening procedures. Two predictors were used; one based on pesticides only, and one for all other compounds (mainly licit and illicit drugs). Both were tested on 30 emerging contaminants commonly found in environmental and wastewater samples by retrospective analysis with QTOF-MS. In addition to help identify the compound of interest, the retention time predictors also allowed over one-third of isobaric chromatographic peaks to be disregarded for further analysis as they were outside the retention time window, thereby enabling the reduction of tedious data processing. This is relevant when applying wide-scope screening for a large number of compounds (e.g. emerging contaminants and pesticides) or when investigating the presence transformation products, where many of the required reference standards are not available at the laboratory. In addition to the retention time prediction, the impact of extracted mass windows was investigated as a complementary tool for the screening. A smaller mass extraction window also removed unwanted peaks from the chromatogram. However, in some cases, even at a narrow mass extraction window of 5 mDa, some isobaric peaks still appeared.

The combination of this simple retention time predictor with extracted mass windows facilitated the removal of many false positives. In this work, 70–80% of compounds studied were able to be found in a ± 2 min window, but a ± 5 min window was needed for $\geq 98\%$ confidence. Our present research is focused on alternate and more sophisticated retention time predictors in order to improve the precision. This would allow the use of narrower time windows, thereby simplifying data processing due to fewer peaks needing to be investigated in the chromatograms.

Acknowledgments

The authors acknowledge the financial support provided by the Plan Nacional de I+D+I, Ministerio de Economía y Competitividad (Project ref CTQ2012-36189) and by Generalitat Valenciana (Group of Excellence Prometeo 2009/054, Prometeo II 2014/023; Collaborative Research on Environment and Food Safety ISIC/2012/016). Richard Bade acknowledges European Union's Seventh Framework Programme for research, technological development and demonstration under Grant Agreement no. [Marie Curie-FP7-PEOPLE Grant no. 317205] for his Early Stage Researcher (ESR) contract.

References

- [1] S.D. Richardson, T.A. Ternes, Water analysis: emerging contaminants and current issues, *Anal. Chem.* 86 (2014) 2813–48. doi:10.1021/ac500508t.
- [2] B. Petrie, E.J. McAdam, M.D. Scrimshaw, J.N. Lester, E. Cartmell, M.E. J., Fate of drugs during wastewater treatment, *TrAC Trends Anal. Chem.* 49 (2013) 145–159. doi:10.1016/j.trac.2013.05.007.
- [3] A. Agüera, M.J. Martínez Bueno, A.R. Fernández-Alba, New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters, *Environ. Sci. Pollut. Res. Int.* 20 (2013) 3496–515. doi:10.1007/s11356-013-1586-0.
- [4] M.J. Gómez, M.M. Gómez-Ramos, O. Malato, M. Mezcua, A.R. Fernández-Alba, Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography-quadrupole-time-of-flight mass spectrometry with an accurate-mass, *J. Chromatogr. A.* 1217 (2010) 7038–54. doi:10.1016/j.chroma.2010.08.070.
- [5] F. Hernández, L. Bijlsma, J. V Sancho, R. Díaz, M. Ibáñez, Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography-quadrupole-time-of-flight-mass spectrometry, *Anal. Chim. Acta.* 684 (2011) 87–97. doi:10.1016/j.aca.2010.10.043.
- [6] A.C. Hogenboom, J.A. van Leerdam, P. de Voogt, Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography-hybrid linear ion trap Orbitrap mass spectrometry, *J. Chromatogr. A.* 1216 (2009) 510–9. doi:10.1016/j.chroma.2008.08.053.
- [7] M. Krauss, H. Singer, J. Hollender, LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns, *Anal. Bioanal. Chem.* 397 (2010) 943–951. doi:10.1007/s00216-010-3608-9.
- [8] F. Hernández, M. Ibáñez, R. Bade, L. Bijlsma, J.V. Sancho, Investigation of pharmaceuticals and illicit drugs in waters by liquid chromatography-high-resolution mass spectrometry, *TrAC Trends Anal. Chem.* 63(2014)140-157. doi:10.1016/j.trac.2014.08.003.
- [9] S. Kern, K. Fenner, H.P. Singer, R.P. Schwarzenbach, J. Hollender, Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry, *Environ. Sci. Technol.* 43 (2009) 7039–7046. doi:10.1021/es901979h.

- [10] R. Díaz, M. Ibáñez, J. V Sancho, F. Hernández, Target and non-target screening strategies for organic contaminants, residues and illicit substances in food, environmental and human biological samples by UHPLC-QTOF-MS, *Anal. Methods*. 4 (2012) 196–209. doi:10.1039/c1ay05385j.
- [11] J. Nurmi, J. Pellinen, A.-L. Rantalainen, Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry, *J. Mass Spectrom.* 47 (2012) 303–12. doi:10.1002/jms.2964.
- [12] T.R. Croley, K.D. White, J.H. Callahan, S.M. Musser, The chromatographic role in high resolution mass spectrometry for non-targeted analysis, *J. Am. Soc. Mass Spectrom.* 23 (2012) 1569–78. doi:10.1007/s13361-012-0392-0.
- [13] K. Héberger, Quantitative structure-(chromatographic) retention relationships, *J. Chromatogr. A*. 1158 (2007) 273–305. doi:10.1016/j.chroma.2007.03.108.
- [14] C. Giaginis, A. Tsantili-Kakoulidou, Quantitative Structure–Retention Relationships as Useful Tool to Characterize Chromatographic Systems and Their Potential to Simulate Biological Processes, *Chromatographia*. 76 (2012) 211–226. doi:10.1007/s10337-012-2374-6.
- [15] D. Livingstone, Theoretical Property Predictions, *Curr. Top. Med. Chem.* 3 (2003) 1171–1192. doi:10.2174/1568026033452078.
- [16] T.H. Miller, A. Musenga, D.A. Cowan, L.P. Barron, Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks., *Anal. Chem.* 85 (2013) 10330–7. doi:10.1021/ac4024878.
- [17] F. Hernández, M. Ibáñez, T. Portolés, M.I. Cervera, J. V Sancho, F.J. López, Advancing towards universal screening for organic pollutants in waters., *J. Hazard. Mater.* 282(2015) 86–95. doi:10.1016/j.jhazmat.2014.08.006.
- [18] M. Ibáñez, L. Bijlsma, A.L.N. van Nuijs, J. V Sancho, G. Haro, A. Covaci, *et al.*, Quadrupole-time-of-flight mass spectrometry screening for synthetic cannabinoids in herbal blends., *J. Mass Spectrom.* 48 (2013) 685–94. doi:10.1002/jms.3217.
- [19] I. V Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, *et al.*, Virtual computational chemistry laboratory - design and description, *J. Comput. Aided. Mol. Des.* 19 (2005) 453–463. doi:10.1007/s10822-005-8694-y.
- [20] C.D.S. Tomlin, ed., *The Pesticide Manual*, Eleventh, British Crop Protection Council, Farnham, 1997.

- [21] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, *et al.*, DrugBank: a knowledgebase for drugs, drug actions and drug targets, *Nucleic Acids Res.* 36 (2008) D901–6. doi:10.1093/nar/gkm958.
- [22] R. Díaz, M. Ibáñez, J. V Sancho, F. Hernández, Building an empirical mass spectra library for screening of organic pollutants by ultra-high-pressure liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 25 (2011) 355–69. doi:10.1002/rcm.4860.
- [23] E. Gracia-Lor, J. V Sancho, F. Hernandez, Multi-class determination of around 50 pharmaceuticals, including 26 antibiotics, in environmental and wastewater samples by ultra-high performance liquid chromatography-tandem mass spectrometry, *J. Chromatogr. a.* 1218 (2011) 2264–2275. doi:10.1016/j.chroma.2011.02.026.
- [24] N. 4/2011 Prescription data: IT del Sistema Nacional de Salud Volumen 35, Subgrupos ATC y Principios activos de mayor consumo en el Sistema Nacional de Salud en 2010, 2011.

Supporting Information

The table “All 595 Compounds used in retention time predictor, empirical retention time and predicted Log Kow” (Table S1), is too large for this thesis, but are available in the online edition of this article: doi:10.1016/j.talanta.2015.02.055. However the other supporting information tables and figures are included here.

Table S2: Predicted retention time, empirical retention time and parent ion ($[M+H]^+$) m/z for all 30 compounds. Those in bold fall outside the retention time window of ± 2 minutes used for the analysis but used in Section 3.3

Type of Compound	Compound	RT (pred)	RT (emp)	$[M+H]^+$
Pesticides	2-hydroxyatrazine	2.66	2.79	198.1355
	2-hydroxyterbuthylazine	3.95	5.76	212.1511
	Atrazine	9.93	9.34	216.1016
	Carbendazine	7.47	3.42	192.0773
	Desethylterbuthylazine	8.47	7.91	202.0859
	Desethylterbumeton	8.95	5.72	198.1355
	Diuron	9.11	8.78	233.0248
	Imazalil	11.20	8.19	297.0561
	Simazine	8.71	8.14	202.0859
	Terbuthylazine	9.99	10.51	230.1172
	Terbutryn	10.72	9.91	242.1439
Pharmaceuticals/ Antibiotics	4-Aminoantipyrine	4.38	3.52	204.1137
	Bezafibrate	10.98	10.86	362.1159
	Carbamazepine	7.93	7.36	237.1028
	Codeine	6.45	2.75	300.1600
	Diclofenac	10.38	12.17	296.0245
	Enrofloxacin	5.56	4.58	360.1723
	Irbesartan	12.60	10.36	429.2403
	Lincomycin	4.93	3.56	407.2216
	Ofloxacin	4.19	4.08	362.1516
	Oxazepam	9.68	9.77	287.0587
	Paracetamol	4.90	2.44	152.0711
	Sulfamethoxazole	5.22	3.91	254.0599
	Temazepam	7.77	10.09	301.0744
	Trimethoprim	5.61	3.64	291.1457
	Valsartan	10.06	10.56	436.2349
Venlafaxine	9.50	7.23	278.2120	
Drugs of Abuse	Benzoylcegonine	2.48	4.53	290.1392
	Cocaine	7.80	4.66	304.1549
	Methamphetamine	7.88	3.05	150.1282

Figure S1: *Correlation of in-house and Waters data*

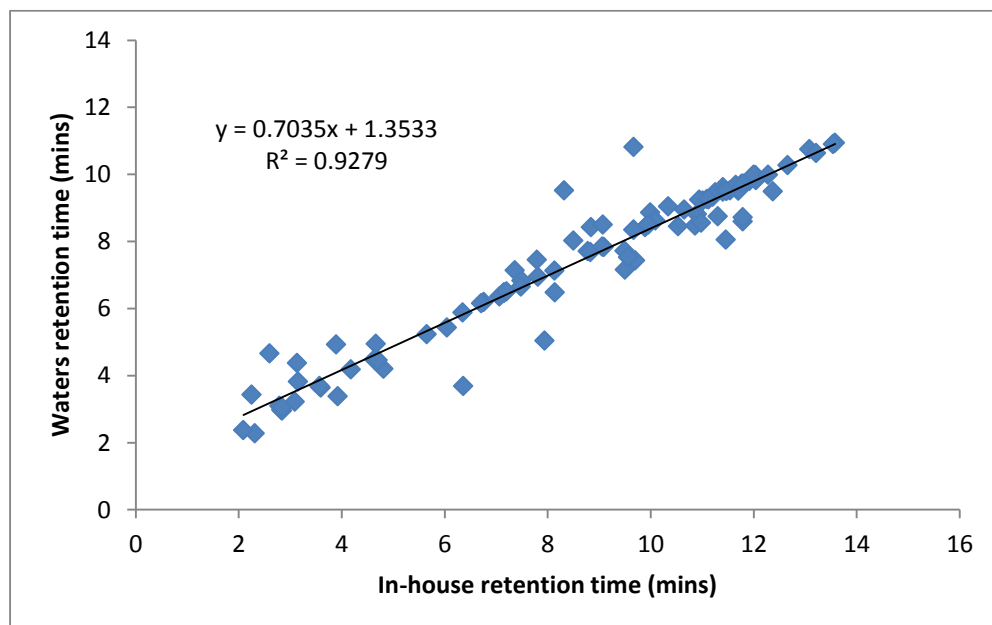
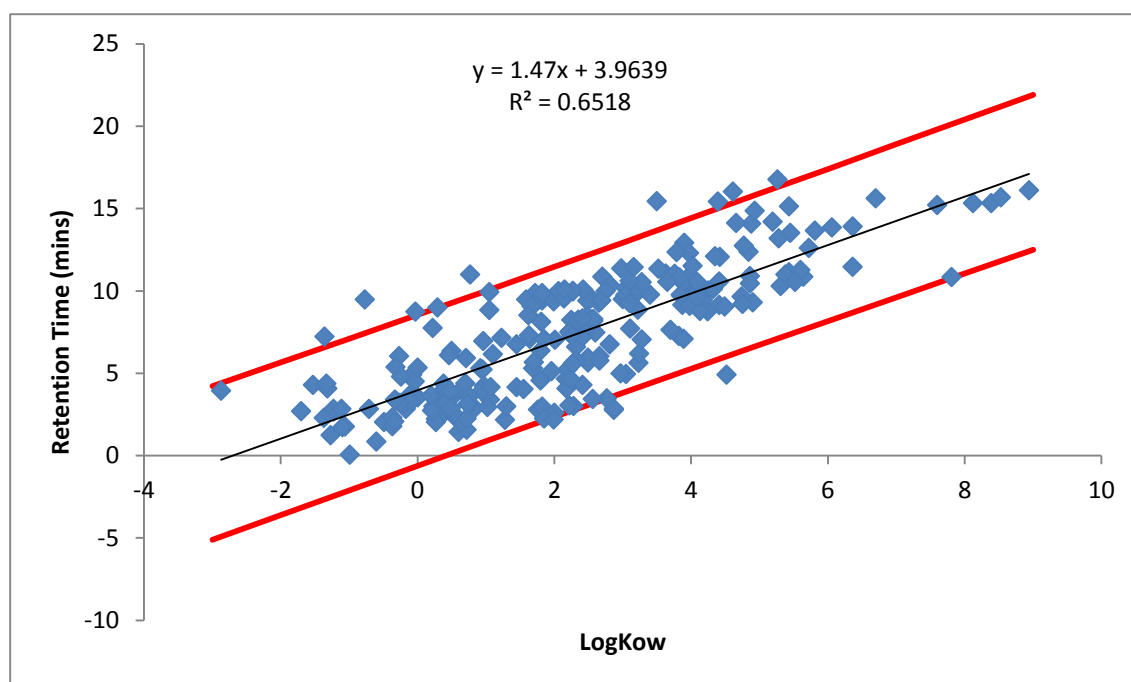


Figure S2: Correlation between retention time and LogKow for all non-pesticides (250 compounds) with 95% confidence interval

3.2.2 Scientific Article 6

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

Richard Bade, Lubertus Bijlsma, Thomas H. Miller, Leon P. Barron, Juan V. Sancho, Felix Hernández

Science of the Total Environment (2015) 538: 934-941



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



Richard Bade^a, Lubertus Bijlsma^a, Thomas H. Miller^b, Leon P. Barron^b,
Juan Vicente Sancho^a, Felix Hernández^{a,*}

^a Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat, E-12071 Castellón, Spain

^b Analytical & Environmental Sciences Division, Faculty of Life Sciences and Medicine, King's College London, 150 Stamford Street, London SE1 9NH, United Kingdom

Abstract

The recent development of broad-scope high resolution mass spectrometry (HRMS) screening methods has resulted in a much improved capability for new compound identification in environmental samples. However, positive identifications at the ng/L concentration level rely on analytical reference standards for chromatographic retention time (t_R) and mass spectral comparisons. Chromatographic t_R prediction can play a role in increasing confidence in suspect screening efforts for new compounds in the environment, especially when standards are not available, but reliable methods are lacking. The current work focuses on the development of artificial neural networks (ANNs) for t_R prediction in gradient reversed-phase liquid chromatography and applied along with HRMS data to suspect screening of wastewater and environmental surface water samples. Based on a compound t_R dataset of > 500 compounds, an optimized 4-layer back-propagation multi-layer perceptron model enabled predictions for 85% of all compounds to within 2 min of their measured t_R for training ($n = 344$) and verification ($n = 100$) datasets. To evaluate the ANN ability for generalization to new data, the model was further tested using 100 randomly selected compounds and revealed 95% prediction accuracy within the 2-minute elution interval. Given the increasing concern on the presence of drug metabolites and other transformation products (TPs) in the aquatic environment, the model was applied along with HRMS data for preliminary identification of pharmaceutically-related compounds in real samples. Examples of compounds where reference standards were subsequently acquired and later confirmed are also presented. To our knowledge, this work presents for the first time, the successful application of an accurate retention time predictor and HRMS data-mining using the largest number of compounds to preliminarily identify new or emerging contaminants in wastewater and surface waters.

Keywords: Retention time prediction; Artificial neural networks; Time-of-flight high resolution mass spectrometry; Screening of emerging contaminants

1. Introduction

The number of emerging contaminants in the aquatic environment is increasing, due to urbanization and subsequent societal and industrial needs (Pal et al., 2014). The development of liquid chromatography–high resolution mass spectrometry (LC–HRMS) technologies has revolutionized the analysis of emerging contaminants in environmental waters, and especially for screening of large numbers of compounds (Agüera et al., 2013, Gómez et al., 2010, Hernández et al., 2011 and Hogenboom et al., 2009). HRMS instruments allow the recording of full-scan spectra with high mass accuracy and resolution, thus making it possible to search for any given compound based on its exact mass.

There has been much interest in improving the confidence in the identification of small molecules with HRMS; from potential positives through to detection and finally confirmation (Hernández et al., 2015a and Schymanski et al., 2014). The main distinguishing factor between these levels is the (non-) availability of reference standards. Suspect screening refers to compounds tentatively identified based solely on HRMS data and comparable spectral libraries. Confirmation requires reference standards. An additional tool to increase the confidence in the tentative identification of compounds for which standards are unavailable is reliable and accurate t_R prediction. This is of particular relevance in the case of degradation/transformation products (TPs), which can reach the aquatic environment in high concentrations, but commonly for which reference standards are less accessible. Chemical risk assessment is therefore significantly challenging for such compounds.

Prediction of t_R plays an important role in the qualitative identification of emerging contaminants. Many different approaches to t_R prediction exist and range from the simple (Kern et al., 2009 and Nurmi et al., 2012) to the complex (Goryński et al., 2013, Ji et al., 2009, Kaliszan et al., 2003 and Ukić et al., 2014a). For example, logKow models can be derived using freely accessible data from chemical databases such as ChemSpider and PubChem, as well as freeware prediction sources such as VCCLABS. Its use in t_R prediction is extremely simple to implement. It is frequently used in environmental studies for the description of the fate of various pollutants and as a simple t_R predictor for TPs (Kern et al., 2009) and emerging contaminants (Bade et al., 2015 and Nurmi et al., 2012). Alongside simple algorithms, other and more complex *in silico* approaches now exist which are based on quantitative structure–retention relationship (QSRR) modeling, including artificial neural networks, support vector machines and random forests (Giaginis and Tsantili-Kakoulidou, 2012 and Héberger, 2007). The principal aim of QSRR is to predict retention data from the molecular structure and its physicochemical properties, using a range of input descriptors and measured t_R data. One QSRR method gaining recent attention for broad screening using high resolution techniques is the use of artificial neural networks (ANNs), a predictive computing technique that has shown itself as a promising t_R predictor with potentially higher accuracy than classical models (Miller et al., 2013 and Ukić et al., 2014b). The design of ANNs were inspired by the human brain and differ from classical computer programs in that they generally employ non-linear learning techniques

using a set of case examples (i.e. a training dataset) (Kaliszan et al., 2003). In the training phase, the ANN requires a range of suitable molecular descriptors as well as the true output value (in this case, measured t_R) to use for comparison with predicted values. At the same time, a second dataset of case examples is often used for verification and to assess overall ANN predictive error. The true output values in the verification set are generally not employed for learning, but the number of training cycles can be stopped by the user or the software when the overall measured error across all cases is at its minimum. Therefore, ANN learning is generally an iterative process and once an acceptable number of training cycles is reached, the optimized ANN can be applied to predict the output where experimentally derived data are unavailable (Miller et al., 2013). In some cases, a third dataset can be used after the model has been finalized to ‘blind test’ the predictive power of the network. Its use is even more pertinent for analyses where large number of new analytes are expected to occur and with potentially high variance from sample to sample, such as in environmental and municipal water samples. Therefore, since information from the sample includes chromatographic t_R as well as HRMS data, it makes this interpretation of suspect occurrence more accessible in the first instance.

The aim of this work was to develop and evaluate ANN for predictions of unknown chromatographic t_R in suspect screening of environmental waters. To the best of our knowledge, this method includes the largest range of physicochemically diverse compounds for this purpose ($n = 544$ in total) and includes both neutral and charged compounds eluted under gradient reversed-phase LC conditions. Lastly, this work aimed to improve upon a recent logKow-based t_R prediction approach (Bade et al., 2015) using the ANNs as an alternative. This work, for the first time, presents the use of ANN for identification of additional suspect compounds (including metabolites and TPs) in wastewater and surface water samples both with and without reference standards.

2. Experimental

2.1. Reagents and chemicals

A total of 544 analytical grade reference materials were used for preparation of model solutions at 25 $\mu\text{g/L}$ or 50 $\mu\text{g/L}$ (diluted from mixed standard solutions in methanol or acetonitrile with water) for ANN modeling of t_R . These included pesticides, drugs of abuse, human/veterinary pharmaceuticals and mycotoxins (See Supplementary Information (SI) Table S1 for all compounds used in this study). These covered a large range of molecular hydrophobicity ($\log K_{ow} - 3$ to 9). Information relating to 595 standards was available (Bade et al., 2015), however after transforming the compounds using SMILES codes, some errors were observed, leading to incomplete data, and a further 42 were removed from the initial ANN method development (Section 3.1) to use in a subsequent blind test (Section 3.2). Further details relating to these compounds can be found elsewhere (Bade et al., 2015 and Hernández et al., 2015b).

2.2. Water samples for suspect identification

A total of 44 composite (24-h) influent and effluent wastewater (IWW and EWW) samples and grab surface water (SW) samples were used to demonstrate the application of the developed ANN model. All these samples were previously used in different studies performed at our lab using the same analytical instrumentation for analysis (Hernández et al., 2015a). All measured t_R data herein were generated using ultra-high pressure liquid chromatography coupled to quadrupole-time of flight mass spectrometry (UHPLC-QTOF-MS).

2.3. UHPLC-QTOF MS

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to a hybrid quadrupole-orthogonal acceleration-TOF mass spectrometer (XEVO G2 QTOF, Waters Micromass, Manchester, UK), using a electrospray ionization (ESI) Z-Spray interface operating in positive mode. The chromatographic separation was performed using an Acquity UPLC BEH C18 100×2.1 mm, $1.7 \mu\text{m}$ particle size column (Waters) at a flow rate of $300 \mu\text{l}/\text{min}$. Gradient elution was performed using mobile phases of A = H₂O and B = MeOH, both containing 0.01% HCOOH. The initial percentage of B was 10%, which was linearly increased to 90% in 14 min, followed by a 2 min isocratic period and, then, returned to initial conditions during 2 min. The total run time was 18 min. Nitrogen was used as the drying gas and nebulizing gas.

MS data were acquired over an m/z range of 50–1000. A capillary voltage of 0.7 kV and cone voltage of 20 V were used. Collision gas was argon 99.995% (Praxair, Valencia, Spain). The desolvation temperature was set to 600 °C, and the source temperature to 135 °C. The column temperature was set to 40 °C. MS data was acquired in MSE mode, selecting a collision energy of 4 eV for low energy (LE) and a ramp of 15–40 eV for high energy (HE). The LE and HE functions settings were for both a scan time of 0.4 s (Hernández et al., 2011 and Ibáñez et al., 2013).

Processing of MS data was made using ChromaLynx XS application manager (within MassLynx v 4.1; Waters Corporation). The following parameters were used: mass window 0.020 Da (for positive ID ≤ 0.010 Da), peak width at 5% height: 6 s, peak-to-peak baseline noise: 1000 and threshold absolute area 500. When manually searching the data for all peaks in an eXtracted Ion Chromatogram (XIC), a chromatographic peak was thought viable when above an intensity threshold of 3000 counts.

2.4. Molecular description and neural network optimization procedures

Compound logD data (for a mobile phase of pH = 3.2) were generated using Percepta PhysChem Profiler (ACD Laboratories, ON, Canada) and for all other descriptors, Parameter Client freeware was used (Virtual Computational Chemistry Laboratory, Munich, Germany). Canonical simplified molecular line entry system strings (SMILES) were created using ChemSpider freeware (Royal Society of Chemistry, UK) for 544 compounds and from these 16 molecular descriptors (as ANN inputs) were generated including the number of double and triple bonds (nDB or nTB), the number of carbon and oxygen atoms (nC or nO), the number of

4–9 membered rings (nR04–nR09), unsaturation index (UI), hydrophilic factor (Hy), Moriguchi and Ghose–Crippen logP (MlogP and AlogP respectively) as well as with software predicted logKow data (Tetko et al., 2005). Prediction of t_R (as the designated single output) via neural networks was performed using Trajan version 6.0 neural network simulator (Trajan Software Ltd., Lincolnshire, U.K.) and compared with experimentally determined t_R via correlation graphs as well as assessment of residual errors. Table S1 has all compounds and their respective predicted and experimental t_R and ANN subset.

3. Results and discussion

In a previous study, we developed a simple t_R prediction model based on logKow of nearly 600 compounds, predicted using freeware (Bade et al., 2015). This resulted in approximately 70% of all compounds being predicted within 2 min of the measured t_R, and 95% within 4 min. This technique was simple to implement and facilitated the removal of several false positives. However, when investigating unknowns and compounds for which reference standards were unavailable, it was concluded that a more robust, accurate and precise methodology was still needed. In this vein, ANNs were considered as an alternative. Recent work successfully used ANN for a similar purpose, albeit using a much smaller set of compounds of 86 and 166 compounds in either study and focused only on pharmaceuticals (Miller et al., 2013 and Munro et al., 2015). It is unlikely that this fully represents the breadth of alternative compound classes and chemistries potentially occurring simultaneously in environmental waters. However, these models successfully predicted t_R for a range of blind test drug compounds in wastewater and urine to warrant further investigation here using a much larger case dataset.

3.1. Prediction of t_R using Artificial Neural Networks (ANNs)

The molecular descriptors chosen were based on the previous work wherein more than 200 descriptors were evaluated (Miller et al., 2013). As the same type of reversed-phase column and LC system were used in both studies, the same descriptors were hoped to provide similar results. These descriptors were used again to also assess the possibility for transferring the model to another laboratory and to extend the prediction to a much larger set of chemically diverse compounds. Collinearity data for all molecular descriptors and retention time are given in the SI. Higher Pearson correlations were observed for hydrophobicity-based descriptors with retention time as was perhaps expected (maximum R = 0.823 with logKow). Similarly, these descriptors also showed some collinearity with each other (R ≤ 0.889) which prevented strong conclusions to be drawn regarding their relative importance to an ANN model. As molecular descriptors selected were from previous investigations, there is also the possibility that additional descriptors may have had more importance to retention prediction on this system. For example, retention on reversed-phase media is not only dependent on hydrophobic interactions, but also steric and shape effects. Large molecules may not interact with the stationary phase well and thus show reduced retention (Wilson et al., 2002). A simple Pearson correlation was examined for a selection of potentially relevant additional descriptors covering charge states, geometrical, topological and

physicochemical properties. Overall, most of these descriptors showed weak relationships < 0.5 except for the descriptor BTLA96 which showed a negative Pearson coefficient of -0.649 .

Over 100,000 network architectures for each model type were initially investigated for their predictive ability across five different ANN model types including 3- and 4-layer multi-layer perceptrons (MLPs), generalized regression neural networks (GRNNs), radial basis functions (RBFs), linear neural networks and probabilistic neural networks (PNN). For training, 344 cases were used along with 100 cases for both verification and blind testing of network performance. Cases were randomly assigned at the beginning of each network test to prevent any bias from pre-selection. Upon selection of the 'best' model statistics (minimum/maximum values, interquartile ranges, standard deviations, medians and means) were generated to ensure that a fair representation of cases and descriptors were present in each dataset subset (see SI). The diversity of ANN types and architectures tested was balanced against the error generated. For network design and testing the omission of input descriptors was included as an option. One hundred of the best networks (software selected) were retained for further investigation which mainly comprised of MLPs, GRNNs and RBFs. Overall for this separation system, the best correlations of predicted versus experimentally measured t_R were observed using MLPs in comparison to all other types and these correlations were in agreement with previous works despite different compounds being used for training, verification and blind testing (Barron et al., 2009, Miller et al., 2013 and Munro et al., 2015). The finalized network was found to be a 4-layer 16-19-9-1 MLP using all 16 molecular descriptors as inputs (Fig. 1). The source code (in C) for this ANN has been attached in the SI. Reducing the number of descriptors further worsened predictive accuracy of the blind test set in general. This ANN type and architecture was chosen based on the lowest absolute errors (i.e. predicted t_R -measured t_R) in the training, verification and blind test sets across all networks. Therefore, ANN architecture was based on performance and the software designer tool was used to optimize the number and composition of hidden layers. The coefficients of determination (R^2) were between 0.86 and 0.90 between the three sets, which was already a marked improvement on that obtained from our previous study at $R^2 = 0.67$. Furthermore, the root-mean-squared error (RMSE) of the blind set of compounds (1.03 min) is less than half that of our previous work (2.19 min) (Bade et al., 2015).

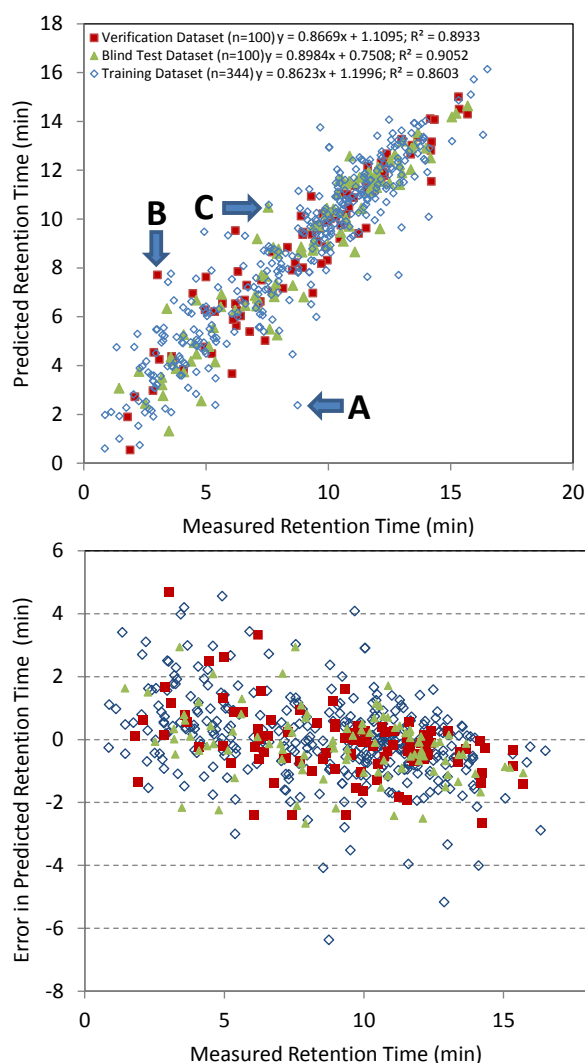


Fig. 1: TOP: Correlation of measured and predicted t_R for all compounds. The worst outliers in each set are also shown: A=atenolol (training set, error of -6.21 minutes); B=levamisole (verification set, +5.19 minutes); C=metosulam (blind set, +3.56 minutes). BOTTOM: Residual errors for all compounds in each dataset. For training, verification and blind test sets, $n=344$, 100 and 100 compounds respectively.

The maximum measured t_R on this chromatographic system was 16.50 min (narasin; an antibiotic) and the lowest was 0.86 min (methamidophos; an organophosphate insecticide). For all compounds within this retention window of 15.64 min, the mean error in t_R prediction was < 6% using this ANN approach for all compounds. Overall, the mean absolute errors and standard deviations were recorded as 0.97 ± 0.95 min (training set); 0.79 ± 0.85 min (verification set); 0.79 ± 0.69 min (blind test set); and 0.91 ± 0.89 min (all sets combined). When focussing specifically on the blind test set which was used to simulate a true application of the approach, 95% of compounds had predicted t_R values within 2.00 min of the measured value and the maximum error was 3.56 min for metosulam (Fig. 1). However, across the other datasets some larger errors were recorded in isolated cases. Table 1 shows that for all datasets, 90% of all 544 compounds could be predicted to within 2.00 min of the measured t_R value. Upon sub-division of the datasets, 85% of the

compounds in the training and verification sets and 95% of compounds in the blind test set were predicted to within two minutes of the measured value. The maximum error recorded within the training set was + 6.25 min (for the beta blocker atenolol; measured t_R = 2.49 min); within the verification set was + 5.19 min (for the anti-helminthic drug levamisole; measured t_R = 3.02 min) and within the blind test set was + 3.56 min (for the pesticide, metosulam; measured t_R = 7.54 min).

Table 1: Summary of predicted t_R errors for all ANN test sets. Numbers given in italics are those which fall below the proposed 2-min window limit

Percentile of compounds	50	55	60	65	70	75	80	85	90	95
Predicted t_R error/min										
All sets (n=544)	<i>0.66</i>	<i>0.75</i>	<i>0.85</i>	<i>0.97</i>	<i>1.05</i>	<i>1.19</i>	<i>1.39</i>	<i>1.56</i>	<i>1.96</i>	2.80
Training set (n=344)	<i>0.70</i>	<i>0.81</i>	<i>0.89</i>	<i>1.02</i>	<i>1.13</i>	<i>1.24</i>	<i>1.45</i>	<i>1.70</i>	<i>2.21</i>	2.87
Verification set (n=100)	<i>0.51</i>	<i>0.59</i>	<i>0.64</i>	<i>0.76</i>	<i>0.89</i>	<i>0.99</i>	<i>1.35</i>	<i>1.44</i>	<i>1.60</i>	2.43
Blind test set (n=100)	<i>0.70</i>	<i>0.78</i>	<i>0.84</i>	<i>0.89</i>	<i>1.01</i>	<i>1.08</i>	<i>1.16</i>	<i>1.34</i>	<i>1.63</i>	<i>1.99</i>

Prediction errors for all cases were investigated again with respect to any apparent trends and it was found that a very slight over-estimation existed for poorly retained compounds, as well as the converse for strongly retained compounds. Recent work focussing on modeling a smaller number of compounds in wastewater also revealed a similarly slight bias, but used a different network type (a GRNN) (Munro et al., 2015). When examining those compounds with absolute errors > 2.00 min (54 compounds in total across all sets), these were spread across the entire compound retention range (mean measured t_R = 7.81 ± 3.93). However, a slight over-estimation was again apparent for 29 compounds with t_R from 1.34–5.38 min (mean measured t_R = 2.97 min). Reduced under-estimation was observed for the remaining 25 compounds eluting between 5.38–16.30 min (mean measured t_R = 10.67 min for these compounds). Seventeen compounds eluting < 5.38 min were over-estimated and t_R for eleven were under-estimated when eluting > 10.00 min.

The contribution of each descriptor towards the final prediction output was investigated using the ANN software sensitivity analysis tool. In this test, each molecular descriptor is removed and treated as missing by the ANN. A new predicted t_R is generated and a ratio calculated between the network error with a given input omitted to the error of the network with a complete input dataset. Ratios > 1 indicated higher importance in the prediction. Perhaps not surprisingly for a reversed-phase chromatography system, the most important molecular descriptors and their measured error ratios were: logD (1.443), logKow (1.182), AlogP (1.114), nO (1.096), UI (1.006) MlogP (1.023), Hy (1.017), nDB (1.012), nR04-nR09 (all 1.000–1.063), nTB (1.004) and nC (1.002) Hydrophobicity-based descriptors are likely to show importance as retention on reversed-phase media is primarily by van der Waals interactions. Again, while these descriptors together show their combined importance to the network, moderate collinearity between them means relative error ratios should be treated with caution for these descriptors (Table SI). However, such collinearity should not adversely affect predictive ability of the model. The lowest ratio was observed for nR04 with a ratio of 1.000 meaning no change in

network performance was measured for its removal. Within the dataset, only 14 compounds had 4-membered rings (amoxicillin, ampicillin, cefaclor, cefadroxil, cefalexin, cefotaxime, cefquinome, cefuroxime, cloxacillin, dicloxacillin, heptenophos, oxacillin, oxasulfuron, and penicillin G). Inclusion of nR04 still resulted in better performance in comparison to any other type or architecture investigated during the network optimization stage and so was retained as a descriptor in the final ANN model. Error ratios discussed above represent that of the entire dataset (training, verification and blind test). Sensitivity analysis (see SI) of each set separately revealed excellent consistency across all sets, showing that predictive accuracy was likely to be the dominant contributors to error ratios rather than over-fitting of the training set alone.

3.2. Use of ANN as a t_R predictor in environmental water samples

In wide-scope screening methods, where thousands of compounds are searched, it is of great importance to have secondary techniques for aid in the identification process. While reference standards can unequivocally confirm the identification of a compound, purchasing and maintaining standards for all compounds is prohibitively expensive. Data acquired from MS fragmentation of suspect compounds may also direct investigations. Furthermore, for many TPs, reference standards are not available and therefore alternate means for confidence in detection/identification are necessary. Although it is not comparatively informative as mass spectra, t_R prediction models can be very helpful in gaining more confidence in the obtained data and reducing time-consuming data processing. Most importantly, the application of t_R prediction is not to replace the use of reference standards, but to help (along with MS/MS data) to direct synthesis efforts for confirmation in the usual manner. Prediction of t_R is best used at the beginning of this process and is especially useful when at a certain exact mass (i.e. (de)protonated molecule of a suspect compound), more than one chromatographic peak appears in the corresponding XIC.

Along these lines, and to test the “blind” skills of the ANN in a real environmental application, the 100 blind compounds as well as an additional set of compounds from our previous study not initially used into the ANN method (a total of 142 compounds), including primarily metabolites and TPs were searched. None of these compounds were in the training or verification sets used for ANN development. From this list of 142 compounds, 46 were finally selected and searched in 44 water samples (EWW, IWW and SW) using ChromaLynx and the ANN predicted t_R , based on their possible occurrence in the environment (Table S2) (Gracia-Lor et al., 2010, Gracia-Lor et al., 2011, Hernández et al., 2015a and Zuccato et al., 2006). For further confidence, and to see how many false positive chromatographic peaks (above the intensity threshold) could be disregarded, fragment ions were also included in the detection process (Table S3). When a compound was identified on the basis of the accurate mass of the (de)protonated molecule and at least one fragment ion in at least one sample it was included in this test, leaving 26 compounds of various chemical classes and including nine metabolites, for only some of which standards were available in our laboratory (Table 2). A t_R window of 2 min was used in this section, as 95% of all compounds in the ANN blind set were within this window, thereby giving high confidence that almost all compounds should be found. The only compound found to have a t_R

outside that of the ANN predicted 2 minute window was codeine. The incorporation of the ANN predicted t_R allowed almost half (49%) of all chromatographic peaks to be ignored, and even after removing codeine from the calculation, 48% could be disregarded. Furthermore, all but three compounds had a reduction in the median number of potential positive peaks, while 11 compounds had a median value of only one chromatographic peak remaining after the introduction of the ANN predicted window.

In this section, examples are shown in the identification of losartan (originally tentatively identified with t_R prediction before a reference standard was purchased) and the tentative identification of the metabolites 10,11-dihydroxy carbamazepine and O-desmethyl venlafaxine (no reference standard available).

Table 2: All compounds used for testing ANN predicted t_R , together with the number of samples each compound was detected, average number of peaks in the XIC and the ± 2 minute window and the predicted and sample t_R

Compound	Detection rate out of 44 samples	Median peaks per XIC (range)	Median peaks inside ± 2 min t_R window (range)	Predicted t_R (min)	Sample t_R (min)	Inaccuracy in predicted t_R (min)
10,11-dihydroxy carbamazepine ^a	30	2 (1-4)	1 (1-3)	6.29	6.66	0.37
2-hydroxy-terbutylazine ^b	6	6 (4-8)	2 (1-3)	4.24	5.50	1.26
4-desmethoxy omeprazole ^a	33	1 (1-2)	1 (1-2)	8.15	6.29	-1.86
4-formylamino-antipyrine ^{b,c}	37	2 (1-8)	2 (1-4)	3.53	3.70	0.17
α -hydroxy metoprolol ^a	17	5 (2-10)	3 (2-4)	2.73	3.30	0.57
Benzoylcegonine ^b	40	3 (1-6)	2 (1-5)	4.11	4.65	0.54
Bezafibrate ^b	2	6 (4-7)	1 (1-1)	10.80	10.78	-0.02
Caffeine ^a	40	3 (1-6)	1 (1-3)	3.17	3.83	0.66
Carbamazepine ^b	40	3 (1-10)	3 (1-5)	7.73	8.82	1.09
Carboxy losartan ^a	35	2 (1-3)	1 (1-1)	10.75	10.44	-0.31
Codeine ^b	25	5 (1-7)	1 (0-3)	4.85	2.46	-2.39
Cotinine ^a	32	10 (4-14)	5 (3-7)	2.00	1.81	-0.19
Diazinon ^{b,c}	4	3 (2-5)	2 (1-2)	11.64	12.50	0.86
Diclofenac ^b	13	1 (1-2)	1 (1-1)	11.71	12.17	0.46
Gemfibrozil ^b	13	4 (3-7)	2 (1-4)	12.29	13.34	1.05
Lidocaine ^a	40	5 (1-11)	2 (1-4)	5.21	4.24	-0.97
Lincomycin ^{b,c}	14	4 (2-8)	2 (1-4)	4.28	3.73	-0.55
Losartan ^b	42	2 (1-4)	2 (1-3)	9.95	10.13	-0.18
Metoprolol ^a	24	3 (1-9)	1 (1-3)	4.07	5.44	1.37
Naproxen ^b	32	7 (4-11)	2 (1-4)	10.06	10.54	0.48
O-desmethylvenlafaxine ^a	29	3 (2-5)	3 (1-5)	5.76	4.68	-1.08
Paraxanthine ^a	39	7 (2-12)	3 (2-4)	2.03	2.97	0.94
Ranitidine ^a	23	4 (2-6)	1 (1-1)	3.03	2.10	-0.93
Terbutylazine ^{b,c}	4	1 (1-3)	1 (1-2)	9.60	10.78	1.18
Trimethoprim ^b	32	7 (2-10)	4 (1-6)	3.11	3.52	0.41
Valsartan ^b	39	2 (1-7)	1 (1-3)	10.90	11.24	0.34

^a: Standard not available in laboratory Sample t_R was based on HRMS data and the incorporation of fragment ions

^b: Standard available in laboratory

^c: Compound in blind set of ANN

3.2.1. Assignment with reference standards

Losartan is a pharmaceutical used to treat hypertension that has been both predicted (Howard and Muir, 2013 and Oosterhuis et al., 2013) and detected in environmental waters (Hernández et al., 2015a and Matsuo et al., 2011). Its presence in a suspect list is thus warranted, and its exact mass was incorporated into our HRMS database. When screening WW and SW samples, XICs at the exact mass of losartan (m/z 423.1700) resulted in two chromatographic peaks (4.63 and 10.13 min) (Fig. 2, bottom left). The ANN t_R predictor was used and calculated a t_R of 9.95 min. This is almost exactly the t_R of one of the peaks, but it is worth noting that even after incorporating a ± 2 minute window only one peak warranted further investigation. Nevertheless, both peaks could have conceivably corresponded to losartan, therefore further research was conducted.

The LE and HE mass spectrum of the peak at 10.13 min was investigated for fragment ions: m/z 207.0917 ($C_{14}H_{11}N_2$, -2.4 ppm), 377.1552 ($C_{22}H_{22}N_4Cl$, $+5$ ppm) and 405.1590 ($C_{22}H_{22}N_4Cl$, -2.3 ppm) (Fig. 2, right). Literature (Hernández et al., 2015a) and the mass spectral database MassBank (Horai et al., 2010) were also searched to aid in the confidence of these fragment ions. It was found that these three fragment ions did indeed correspond to losartan. As seen in the figure, all associated fragments corresponded to the peak at 10.13 min, following the assertion of the t_R predictor. Furthermore, it is clearly seen that none of the fragment ions correspond with that of the other peak in the LE (4.63 min). A standard was later purchased and injected, unequivocally confirming that the peak at 10.13 min did correspond to losartan.

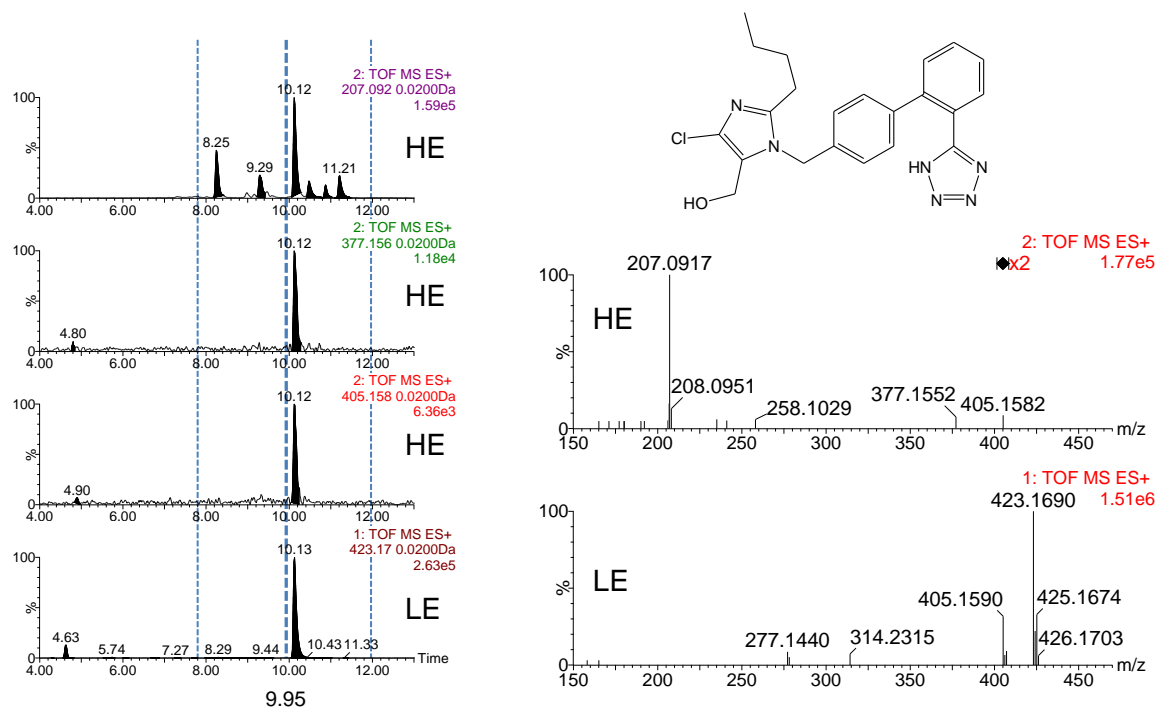


Fig. 2: LEFT: *e*Xtracted Ion Chromatograms (XICs) of losartan in an EWW sample, together with ANN t_R prediction (9.95 min) and associated ± 2 minute window (dotted lines; centre dashed line is the ANN-predicted value). RIGHT: Mass spectra from LE (bottom) and HE (top), showing fragment ions (423.158, 377.156, 207.092).

3.2.2. Tentative identification of metabolites without standard reference materials

The elimination of false positives in suspect analysis is challenging, especially the environmental matrices investigated herein (SW and WW) as thousands of compounds may be present. Suspect screening, by definition, does not rely on reference standards (Krauss et al., 2010). While the exact mass capability of HRMS has gone some way to avoid false positives, even at narrow mass window chromatograms, matrix interferences can be present, thereby hindering confident identification (Bade et al., 2015 and Croley et al., 2012). A precise t_R predictor can therefore be a great additional means of identification.

Most research of emerging contaminants in the environment has focused on parent compounds, however many compounds can be at least partially metabolized or degraded in natural conditions (Jakimska et al., 2014). In this respect, the tentative identification of two major metabolites of carbamazepine and venlafaxine were explored: 10,11-dihydroxy carbamazepine and O-desmethyl venlafaxine (Fig. 3a). More than one chromatographic peak was observed at the exact mass of each protonated molecule. The ANN t_R predictor was then used to try to help minimize the number of peaks to be analyzed, with the predicted t_R calculated to be 6.29 and 5.76 min for 10,11-dihydroxy carbamazepine and O-desmethyl venlafaxine, respectively.

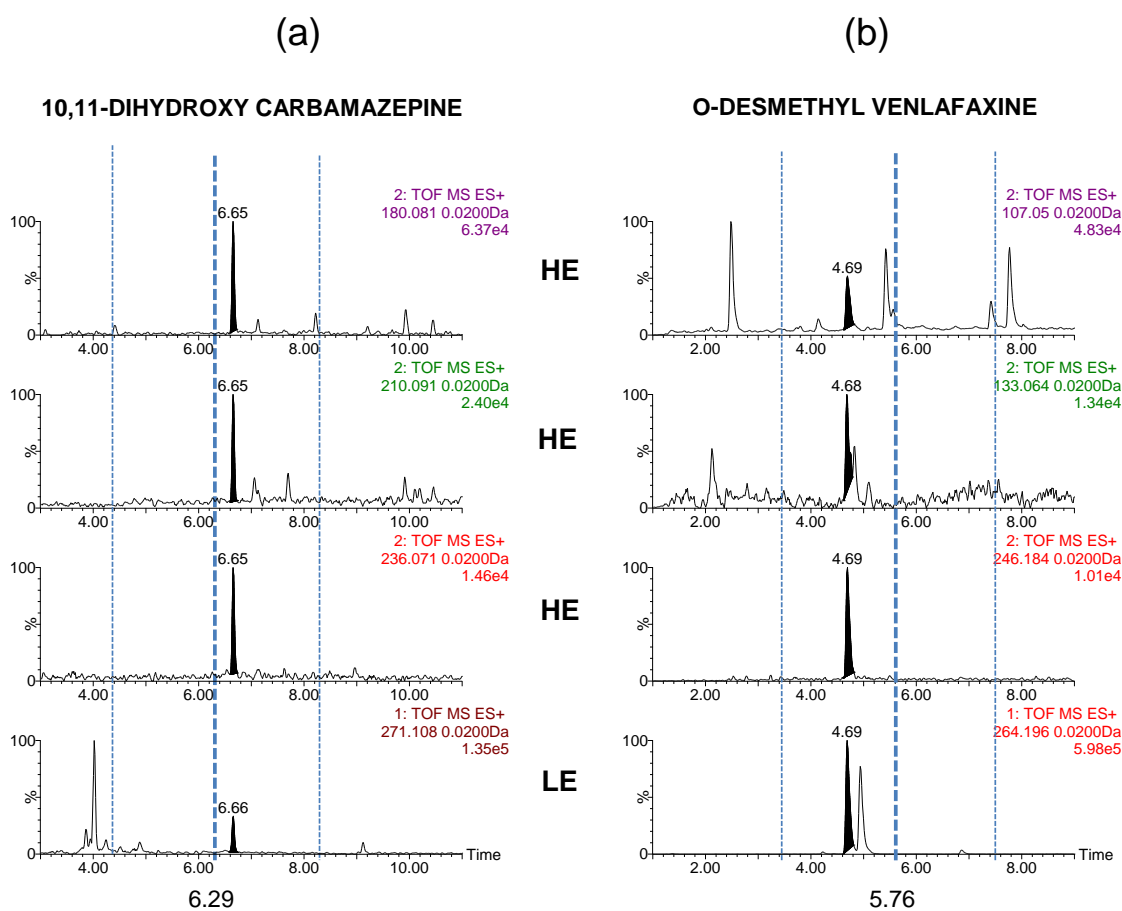


Fig. 3: Tentative identification of 10,11-dihydroxy carbamazepine (a) and O-desmethyl venlafaxine (b) in an IWW and EWW sample respectively, together with ANN prediction and ± 2 minute window.

Two large (~ 4.00 and 6.66 min) and one small (~ 4.90 min) peaks are seen in the LE XIC of 10,11-dihydroxy carbamazepine ($m/z = 271.1080$). The predicted t_R was calculated to be 6.29 min and by including a ± 2 minute window as in Section 3.2.1, the large peak at ~ 4.00 min could be disregarded, leaving the peaks at 6.66 and ~ 4.90 min. The ability to focus on fewer peaks is the primary aim and benefit of t_R prediction in environmental screening applications for unknowns. While having only one (correct) peak remaining is ideal, being able to disregard some peaks gives credence to the use of t_R prediction in the identification process. To aid in the differentiation of these peaks, fragment ions were sought in literature, whereby one group performed an MS/MS experiment with a QTOF instrument to find the fragment ions of 10,11-dihydroxy carbamazepine (271.1080 , 236.0706 , 210.0913 and 180.0808) (Ferrer and Thurman, 2012). As the XICs in the figure show, all fragment ions have the peak at 6.65 min in common. This provides great confidence that this peak is indeed from 10,11-dihydroxy carbamazepine, however for unequivocal identification, a standard would still have to be purchased.

The example of O-desmethyl venlafaxine represented the worst case scenario, where no peaks could be removed after application of the ± 2 minute limit window. However, it must be noted that this was a rare case and only occurred for three of 26 compounds. In the LE XIC, two large (4.69 and 5.00 min) and two small peaks (~ 4.20 and 6.50 min) are seen. Even after incorporating the ANN predicted t_R and associated window (5.76 ± 2 min), all four peaks are still of interest. With the peaks being so close together, even using a ± 1.3 minute window (corresponding to $\sim 80\%$ of compounds being successfully inside this window) would result in the correct elimination of only one small peak. In this situation, fragment ions have to be used to gain further information on the identity. Fragment ions were thus searched in literature (Herrera-Lopez et al., 2014) and investigated in the HE. Fig. 3b shows all ions corresponding to the peak at 4.69 min. While the ANN predicted t_R and associated window did include O-desmethyl venlafaxine, all other peaks in the LE XIC were also inside, meaning that no further confidence could be gained by t_R prediction, rather through the investigation of fragment ions. Nevertheless, the combination of ANN predicted t_R and fragment ions led to the tentative identification of this compound.

While the examples explained here show the successful assignment of chromatographic peaks, it is impossible to have total confidence with t_R prediction. In cases where there is more than one peak in the XIC, and the predicted peak is found to be incorrect, the peaks slightly outside the prediction window will also have to be investigated. Nevertheless, with data processing nowadays being the most time consuming part of environmental screening methods, the time saved by incorporating t_R prediction outweighs the possibility of false negatives (and positives).

These examples clearly show the utility of ANN as a t_R predictor, not just for its ability to disregard some false positive peaks, but also for its accuracy and subsequent confidence for tentatively identified compounds. It is therefore recommended when performing large scope (e.g. > 1000 compounds) screening of environmental

samples to include accurate t_R prediction in the strategy used for identification. This is particularly useful in the investigation of metabolites and TPs, for which standards can be very costly or unavailable.

4. Conclusions

This work showed the development and use of a t_R predictor based on artificial neural networks. In particular, a four layer multilayer perceptron successfully modeled retention of 544 compounds under these separation conditions. Overall, 90% of all compounds eluted within 2.00 min of the predicted value and for 100 blind test compounds, 95% were predicted within this window. The network was applied to additional suspect compound occurrence in wide-scope screening based on the use of LC-HRMS, demonstrating that it can reduce the number of false negatives or positives. This saves time and effort in the tentative identification of the compounds detected, as only those chromatographic peaks that fit the predicted t_R need to be focused on. Several representative examples are given to illustrate the usefulness of the complementary use of precise t_R prediction in large suspect screening of emerging contaminants. It is recommended to include this prediction for the identification of suspect compounds, particularly in the investigation of metabolites and TPs of organic contaminants, for which reference standards are commonly less accessible.

Acknowledgments

Richard Bade acknowledges the European Union for his Early Stage Researcher (ESR) contract as part of the EU-International Training Network SEWPROF (Marie Curie — PEOPLE Grant #317205).

The financial support from the Spanish Ministry of Economy and Competitiveness (Ref CTQ2012-36189), and Generalitat Valenciana, Spain (research group of excellence PROMETEO II/2014/023; ISIC 2012/016) is also acknowledged.

Thanks are also extended to the Biotechnology and Biological Sciences Research Council (BBSRC) and AstraZeneca (under the Global SHE research program) for award of a CASE industrial scholarship for T. Miller who collaborated on this project (Reference: BB/K501177/1).

References

- Agüera et al., 2013: A. Agüera, M.J. Martínez Bueno, A.R. Fernández-Alba New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters *Environ. Sci. Pollut. Res. Int.*, 20 (2013), pp. 3496–3515 <http://dx.doi.org/10.1007/s11356-013-1586-0>
- Bade et al., 2015: R. Bade, L. Bijlsma, J.V. Sancho, F. Hernández Critical evaluation of a simple retention time predictor based on LogK_{ow} as a complementary tool in the identification of emerging contaminants in water *Talanta*, 139 (2015), pp. 143–149 <http://dx.doi.org/10.1016/j.talanta.2015.02.055>
- Barron et al., 2009: L. Barron, J. Havel, M. Purcell, M. Szpak, B. Kelleher, B. Paull Predicting sorption of pharmaceuticals and personal care products onto soil and digested sludge using artificial neural networks *Analyst*, 134 (2009), pp. 663–670 <http://dx.doi.org/10.1039/b817822d>
- Croley et al., 2012: T.R. Croley, K.D. White, J.H. Callahan, S.M. Musser The chromatographic role in high resolution mass spectrometry for non-targeted analysis *J. Am. Soc. Mass Spectrom.*, 23 (2012), pp. 1569–1578 <http://dx.doi.org/10.1007/s13361-012-0392-0>
- Ferrer and Thurman, 2012: I. Ferrer, E.M. Thurman Analysis of 100 pharmaceuticals and their degradates in water samples by liquid chromatography/quadrupole time-of-flight mass spectrometry *J. Chromatogr. A*, 1259 (2012), pp. 148–157 <http://dx.doi.org/10.1016/j.chroma.2012.03.059>
- Giaginis and Tsantili-Kakoulidou, 2012: C. Giaginis, A. Tsantili-Kakoulidou Quantitative structure–retention relationships as useful tool to characterize chromatographic systems and their potential to simulate biological processes *Chromatographia*, 76 (2012), pp. 211–226 <http://dx.doi.org/10.1007/s10337-012-2374-6>
- Gómez et al., 2010: M.J. Gómez, M.M. Gómez-Ramos, O. Malato, M. Mezcuca, A.R. Fernández-Alba Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography–quadrupole-time-of-flight mass spectrometry with an accurate-mass *J. Chromatogr. A*, 1217 (2010), pp. 7038–7054 <http://dx.doi.org/10.1016/j.chroma.2010.08.070>
- Goryński et al., 2013: K. Goryński, B. Bojko, A. Nowaczyk, A. Buciński, J. Pawliszyn, R. Kaliszan Quantitative structure–retention relationships models for prediction of high performance liquid chromatography retention time of small molecules: endogenous metabolites and banned compounds *Anal. Chim. Acta*, 797 (2013), pp. 13–19 <http://dx.doi.org/10.1016/j.aca.2013.08.025>
- Gracia-Lor et al., 2010: E. Gracia-Lor, J.V. Sancho, F. Hernandez Simultaneous determination of acidic, neutral and basic pharmaceuticals in urban wastewater by ultra high-pressure liquid chromatography–tandem mass spectrometry *J. Chromatogr. A*, 1217 (2010), pp. 622–632 <http://dx.doi.org/10.1016/j.chroma.2009.11.090>

Gracia-Lor et al., 2011: E. Gracia-Lor, J.V. Sancho, F. Hernandez Multi-class determination of around 50 pharmaceuticals, including 26 antibiotics, in environmental and wastewater samples by ultra-high performance liquid chromatography–tandem mass spectrometry *J. Chromatogr. A*, 1218 (2011), pp. 2264–2275 <http://dx.doi.org/10.1016/j.chroma.2011.02.026>

Héberger, 2007: K. Héberger Quantitative structure–(chromatographic) retention relationships *J. Chromatogr. A*, 1158 (2007), pp. 273–305 <http://dx.doi.org/10.1016/j.chroma.2007.03.108>

Hernández et al., 2011: F. Hernández, L. Bijlsma, J.V. Sancho, R. Díaz, M. Ibáñez Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography–quadrupole-time-of-flight-mass spectrometry *Anal. Chim. Acta*, 684 (2011), pp. 87–97 <http://dx.doi.org/10.1016/j.aca.2010.10.043>

Hernández et al., 2015a: F. Hernández, M. Ibáñez, A.-M. Botero-Coy, R. Bade, M.C. Bustos-López, J. Rincón, A. Moncayo, L. Bijlsma LC-QTOF MS screening of more than 1,000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá Colombia. *Anal. Bioanal. Chem.*, 407 (2015), pp. 6405–6416 <http://dx.doi.org/10.1007/s00216-015-8796-x>

Hernández et al., 2015b: F. Hernández, M. Ibáñez, T. Portolés, M.I. Cervera, J.V. Sancho, F.J. López Advancing towards universal screening for organic pollutants in waters *J. Hazard. Mater.*, 282 (2015), pp. 86–95 <http://dx.doi.org/10.1016/j.jhazmat.2014.08.006>

Herrera-Lopez et al., 2014: S. Herrera-Lopez, M.D. Hernando, E. García-Calvo, A.R. Fernández-Alba, M.M. Ulaszewska Simultaneous screening of targeted and non-targeted contaminants using an LC–QTOF-MS system and automated MS/MS library searching *J. Mass Spectrom.*, 49 (2014), pp. 878–893 <http://dx.doi.org/10.1002/jms.3428>

Hogenboom et al., 2009: A.C. Hogenboom, J.A. van Leerdam, P. de Voogt Accurate mass screening and identification of emerging contaminants in environmental samples by liquid chromatography-hybrid linear ion trap Orbitrap mass spectrometry *J. Chromatogr. A*, 1216 (2009), pp. 510–519 <http://dx.doi.org/10.1016/j.chroma.2008.08.053>

Horai et al., 2010: H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M.Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka MassBank: a public repository for sharing mass spectral data for life sciences *J. Mass Spectrom.*, 45 (2010), pp. 703–714 <http://dx.doi.org/10.1002/jms.1777>

Howard and Muir, 2013: P.H. Howard, D.C.G. Muir Identifying new persistent and bioaccumulative organics among chemicals in commerce. III: byproducts, impurities, and transformation products *Environ. Sci. Technol.*, 47 (2013), pp. 5259–5266 <http://dx.doi.org/10.1021/es4004075>

Ibáñez et al., 2013: M. Ibáñez, L. Bijlsma, A.L.N. van Nuijs, J.V. Sancho, G. Haro, A. Covaci, F. Hernández Quadrupole-time-of-flight mass spectrometry screening for synthetic cannabinoids in herbal blends *J. Mass Spectrom.*, 48 (2013), pp. 685–694 <http://dx.doi.org/10.1002/jms.3217>

Jakimska et al., 2014: A. Jakimska, A. Kot-Wasik, J. Namieśnik The current state-of-the-art in the determination of pharmaceutical residues in environmental matrices using hyphenated techniques *Crit. Rev. Anal. Chem.*, 44 (2014), pp. 277–298 <http://dx.doi.org/10.1080/10408347.2013.835244>

Ji et al., 2009: C. Ji, Y. Li, L. Su, X. Zhang, X. Chen Quantitative structure–retention relationships for mycotoxins and fungal metabolites in LC–MS/MS *J. Sep. Sci.*, 32 (2009), pp. 3967–3979 <http://dx.doi.org/10.1002/jssc.200900441>

Kaliszan et al., 2003: R. Kaliszan, T. Baczek, A. Bucinĳnski, B. Buszewski, M. Sztupecka Prediction of gradient retention from the linear solvent strength (LSS) model, quantitative structure–retention relationships (QSRR), and artificial neural networks (ANN) *J. Sep. Sci.*, 26 (2003), pp. 271–282 <http://dx.doi.org/10.1002/jssc.200390033>

Kern et al., 2009: S. Kern, K. Fenner, H.P. Singer, R.P. Schwarzenbach, J. Hollender Identification of transformation products of organic contaminants in natural waters by computer-aided prediction and high-resolution mass spectrometry *Environ. Sci. Technol.*, 43 (2009), pp. 7039–7046 <http://dx.doi.org/10.1021/es901979h>

Krauss et al., 2010: M. Krauss, H. Singer, J. Hollender LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns *Anal. Bioanal. Chem.*, 397 (2010), pp. 943–951 <http://dx.doi.org/10.1007/s00216-010-3608-9>

Matsuo et al., 2011: H. Matsuo, H. Sakamoto, K. Arizono, R. Shinohara Behavior of pharmaceuticals in waste water treatment plant in Japan *Bull. Environ. Contam. Toxicol.*, 87 (2011), pp. 31–35 <http://dx.doi.org/10.1007/s00128-011-0299-7>

Miller et al., 2013: T.H. Miller, A. Musenga, D.A. Cowan, L.P. Barron Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks *Anal. Chem.*, 85 (2013), pp. 10330–10337 <http://dx.doi.org/10.1021/ac4024878>

Munro et al., 2015: K. Munro, T.H. Miller, C.P.B. Martins, A.M. Edge, D. a Cowan, L.P. Barron Artificial neural network modelling of pharmaceutical residue retention times in wastewater extracts using gradient liquid

chromatography–high resolution mass spectrometry data *J. Chromatogr. A*, 1396 (2015), pp. 33–44
<http://dx.doi.org/10.1016/j.chroma.2015.03.063>

Nurmi et al., 2012: J. Nurmi, J. Pellinen, A.-L. Rantalainen Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry *J. Mass Spectrom.*, 47 (2012), pp. 303–312 <http://dx.doi.org/10.1002/jms.2964>

Oosterhuis et al., 2013: M. Oosterhuis, F. Sacher, T.L. ter Laak Prediction of concentration levels of metformin and other high consumption pharmaceuticals in wastewater and regional surface water based on sales data *Sci. Total Environ.*, 442 (2013), pp. 380–388 <http://dx.doi.org/10.1016/j.scitotenv.2012.10.046>

Pal et al., 2014: A. Pal, Y. He, M. Jekel, M. Reinhard, K.Y.H. Gin Emerging contaminants of public health significance as water quality indicator compounds in the urban water cycle *Environ. Int.*, 71 (2014), pp. 46–62
<http://dx.doi.org/10.1016/j.envint.2014.05.025>

Schymanski et al., 2014: E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender Identifying small molecules via high resolution mass spectrometry: communicating confidence *Environ. Sci. Technol.* (2014) <http://dx.doi.org/10.1021/es5002105>

Tetko et al., 2005: I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. Palyulin, E. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko Virtual computational chemistry laboratory — design and description *J. Comput. Aided Mol. Des.*, 19 (2005), pp. 453–463
<http://dx.doi.org/10.1007/s10822-005-8694-y>

Ukić et al., 2014a: Š. Ukić, M. Novak, P. Žuvela, N. Avdalović, Y. Liu, B. Buszewski, T. Bolanča Development of gradient retention model in ion chromatography. Part I: conventional QSRR approach *Chromatographia*, 77 (2014), pp. 985–996 <http://dx.doi.org/10.1007/s10337-014-2653-5>

Ukić et al., 2014b: Š. Ukić, M. Novak, P. Žuvela, N. Avdalović, Y. Liu, B. Buszewski, T. Bolanča Development of gradient retention model in ion chromatography. Part II: artificial intelligence QSRR approach *Chromatographia*, 77 (2014), pp. 997–1007 <http://dx.doi.org/10.1007/s10337-014-2653-5>

Wilson et al., 2002: N.S. Wilson, J.W. Dolan, L.R. Snyder, P.W. Carr, L.C. Sander Column selectivity in reversed-phase liquid chromatography: III. The physico-chemical basis of selectivity *J. Chromatogr. A*, 961 (2002), pp. 217–236 [http://dx.doi.org/10.1016/S0021-9673\(02\)00658-1](http://dx.doi.org/10.1016/S0021-9673(02)00658-1)

Zuccato et al., 2006: E. Zuccato, S. Castiglioni, R. Fanelli, G. Reitano, R. Bagnati, C. Chiabrando, F. Pomati, C. Rossetti, D. Calamari Pharmaceuticals in the environment in Italy: causes, occurrence, effects and control *Environ. Sci. Pollut. Res. Int.*, 13 (2006), pp. 15–21 <http://dx.doi.org/10.1065/espr2006.01.004>

Supporting Information

The tables in the supporting information: A All compounds used for the initial ANN predictions, including experimental t_R , predicted t_R , absolute error and subset (Table S1), All 46 compounds investigated in the samples, including predicted t_R and exact mass (m/z) of their protonated molecule (Table S2), All compounds searched in the secondary blind test, together with exact mass (m/z) of protonated molecules and fragment ions (Table S3) as well as information relating to the descriptors used, collinearity tests and ANN sensitivity analysis are too large for this thesis, but are available in the online version: doi: 10.1016/j.scitotenv.2015.08.078.

3.3 Discussion of Results

These two scientific articles have shown the potential of t_R prediction: from an initial, very simple approach based solely on $\log K_{ow}$ to a more complex, computational, QSRR method based on ANN.

The investigation into t_R prediction was begun because of the ever-increasing number of compounds able to be detected based on high-resolution mass spectrometry in our laboratory. With the expense associated with purchasing reference standards and the expiration date of stock solutions, a complementary means of tentative identification would give more confidence to detected peaks before having to purchase standards. Moreover, for degradation and metabolism studies, where reference standards are oftentimes impossible to obtain, a secondary means of confidence would be of great use.

Initially, a simple approach was advised, which any laboratory could also do without having to buy expensive software: the $\log K_{ow}$ only approach. This approach was indeed very simple, from a computational point of view, being based on a single descriptor and the retention times of all the reference standards we had previously injected into our UPLC-QTOF MS as well as a set from Waters, which had been run on the same instrument using a slightly different gradient. $\log K_{ow}$ was decided upon as the descriptor of choice as it mimics the processes involved in reversed phase as well as being known to have a good correlation with t_R for C_{18} columns with methanol as the organic modifier. This was shown to be the case, however with almost 600 compounds from differing chemical families and physicochemical properties, an R^2 of slightly under 0.70 was the best that could be achieved. This imperfect correlation meant that the t_R window would have to either be made wide (i.e. ± 4 minutes) to have a statistically significant 90-95% of the compounds being found or be made small enough to be analytically relevant (i.e. ± 2 minutes) with the related prediction error leading to 70-80% of all compounds being found. The latter was chosen, with the associated disadvantages described in **Scientific Article 5**. While this approach did succeed in its initial goal to produce a simple and free t_R predictor, a more precise approach was sought.

An approach was made to Leon Barron at King's College, London, who had successfully applied ANN for t_R prediction for anti-doping screening of urine, for a collaboration using our compilation of almost 600 reference standards. It turned out that using the same fifteen descriptors they had used (Miller et al., 2013) as well as $\log K_{ow}$ led to a very good prediction, with up to 95% of all compounds found within two minutes of their experimental t_R . As shown in **Scientific Article 6** this enabled its

use in the tentative identification of compounds, metabolites and transformation products for which reference standards were unavailable.

To compare the two approaches, **Figure 3.1** shows the use of both t_R prediction methods in the detection of three common emerging contaminants: bezafibrate (a), trimethoprim (b) and codeine (c), using the t_R window of ± 2 minutes as in both scientific articles. These three examples portray three specific cases: when the compound is within the two minute window of the simple $\log K_{ow}$ only approach (a); when the compound is outside the window of the $\log K_{ow}$ only but inside that of the ANN approach (b); and finally where the compound is neither within the $\log K_{ow}$ only nor ANN windows (c).

The first example (a) shows the utility of the simple $\log K_{ow}$ only approach, where the predicted t_R (± 2 minutes) includes the peak of bezafibrate (10.91), whilst also enabling all other chromatographic peaks to be disregarded. The second example (b) is one typical of suspect screening, with more than one large peak inside the XIC. While the $\log K_{ow}$ prediction is unable to remove (or include) either of the large peaks, the ANN prediction includes the peak of benzoylecgonine. The final example is very uncommon and unwanted: where both approaches fail. The t_R prediction windows do not include any of the three peaks of the XIC, where the predictions from $\log K_{ow}$ (6.46 mins) and ANN (4.85 mins) are nowhere near the three peaks in the XIC, let alone the peak of codeine (2.46 mins). This final case is one which, albeit rare, can happen; as explained in both scientific articles, whereby not 100% of all compounds can be found within two minutes. In order to improve the chance of the compounds being within the t_R window, it may be necessary to widen the window, which can have the adverse effect of increasing the number of potential peaks needing to be later analysed. However, as has been shown, especially for the ANN approach, a window of ± 2 minutes is more than enough for the vast majority of compounds.

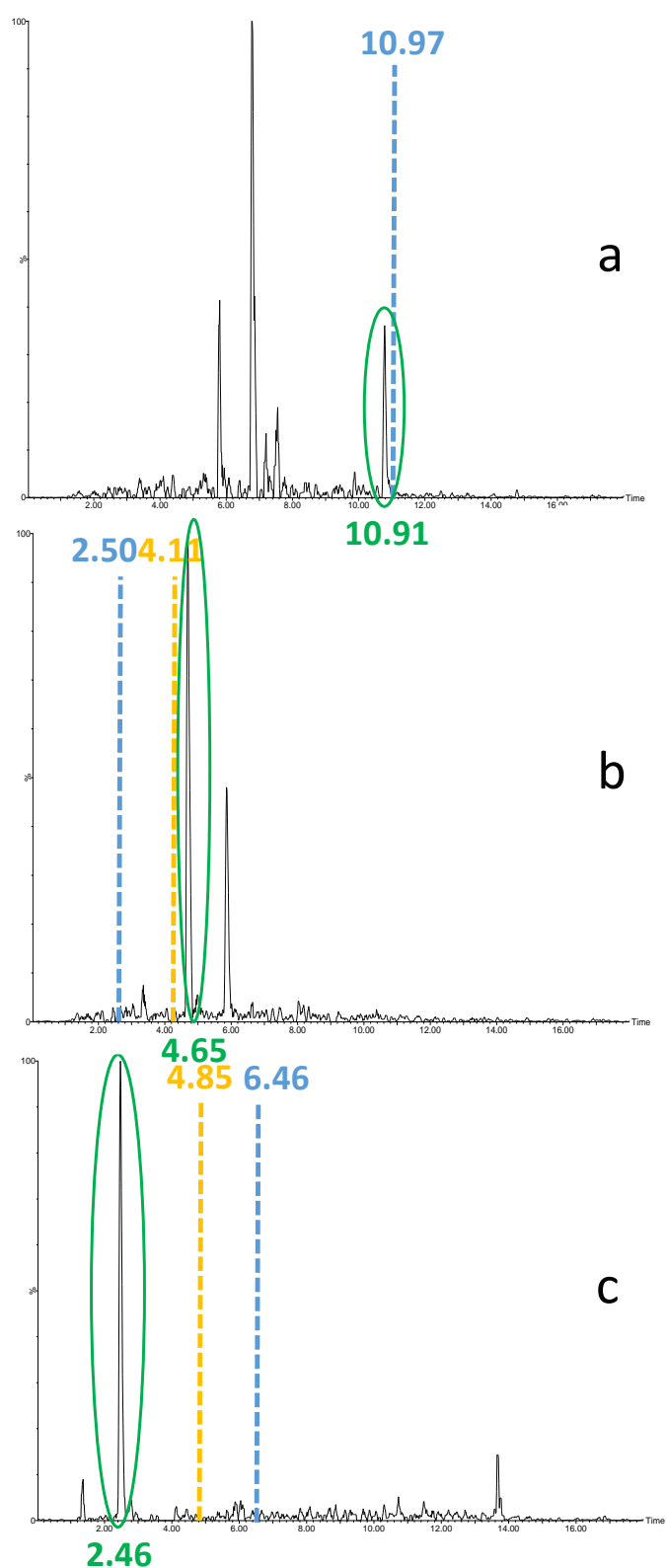


Figure 3.1: Extracted ion chromatograms of three compounds; bezafibrate (a), benzoylecgonine (b) and codeine (c) at a mass extraction window of 10mDa and incorporating the retention time predictions of both the $\log K_{ow}$ only approach (blue), ANNs (yellow), while green is the peak of actual compound.

While the examples of **Figure 3.1** do show the advantages and disadvantages of t_R prediction in terms, all had reference standards, so it was straightforward to finally identify the compounds. However, when investigating new compounds, metabolites and transformation products, for which standards may not be available, it is imperative to have a complementary means of identification; a precise t_R predictor such as ANNs alongside conventional mass spectrometric fragment ion analysis. **Figure 3** in **Scientific Article 6** shows two examples of 10,11-dihydroxy carbamazepine and *O*-desmethyl venlafaxine, the first of which shows the perfect example of the utility of a precise t_R predictor, able to remove three peaks from later analysis, while the latter shows the situation of two almost co-eluting peaks which are unable to be distinguished solely with t_R prediction. A further example is shown in **Figure 3.2**, of carboxy losartan, a metabolite of the antihypertensive, losartan. It is clear to see in the low energy XIC (where solely the parent compound should be seen) there are two peaks, one small one at ~8 minutes and a second, larger one at 10.45 minutes. While some analysts performing suspect screening may (sometimes erroneously) assume that the larger peak is their compound of interest, the introduction of t_R prediction (10.75 mins) immediately removes the smaller peak from further investigation, vindicated by the fragment ions relating solely to the peak at 10.45 mins.

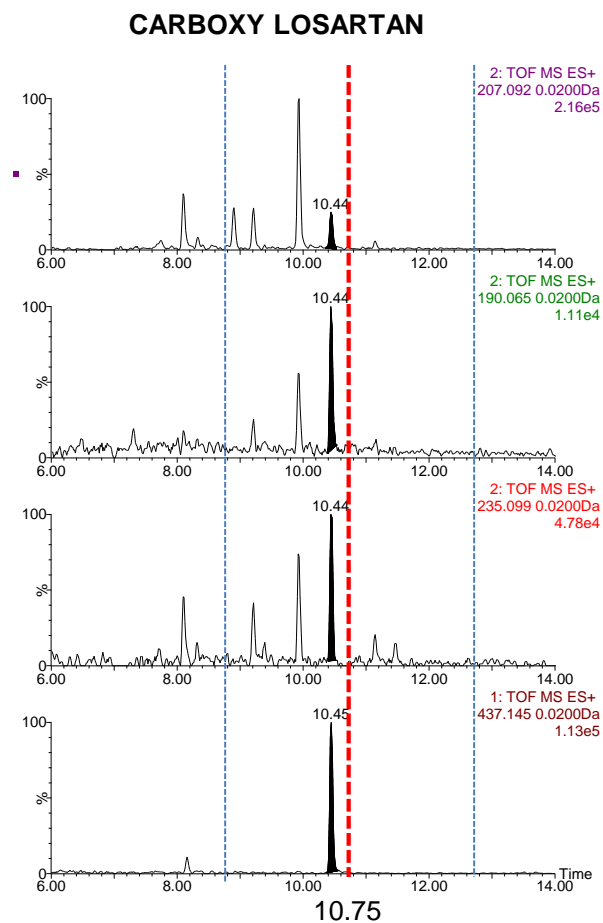


Figure 3.2: Tentative identification of carboxy losartan following the introduction of ANN optimised t_R prediction

The previous examples have shown the way toward tentative identification using t_R prediction as the initial point of call. However, this is not necessarily always the case, with some laboratories performing mass spectral analysis (i.e. common fragmentation pathways) first to determine the fragment ions relating to metabolites of a known parent compound. This can provide some idea of which chromatographic peaks give rise to which metabolite. **Figure 3.3** shows the common fragmentation pathway route for the tentative identification of six metabolites of the antihypertensive diltiazem (Ibáñez et al., 2016) with no reference standards being available to aid in their identification. To provide further confidence to their identification, t_R prediction was used. As can be seen, all predictions were within the ± 2 minute window used in previous examples, most within one minute.

One interesting aspect of this diltiazem example is the observance of isomers, which are notoriously difficult to distinguish based solely on mass spectral analysis. Without spending further analysis time and expense in synthesising standards, to help differentiate the isomers, t_R prediction could be used. As t_R prediction is based on the chemical structure, isomers can have slightly different properties,

giving rise to different t_R . While total confirmation cannot be made with t_R prediction, a qualitative idea on the elution order of the isomers is possible. This is shown perfectly where the two sets of isomers: O-deacetyl-O-desmethyldiltiazem and O-deacetyl-N-desmethyldiltiazem and O-desmethyl diltiazem and N-desmethyl diltiazem could not be distinguished based on mass spectral analysis. Although the predicted t_R were similar, the metabolites were able to be assigned based on the elution order (i.e. 6.23, 6.48, 7.06, 7.52 for the aforementioned metabolites, respectively).

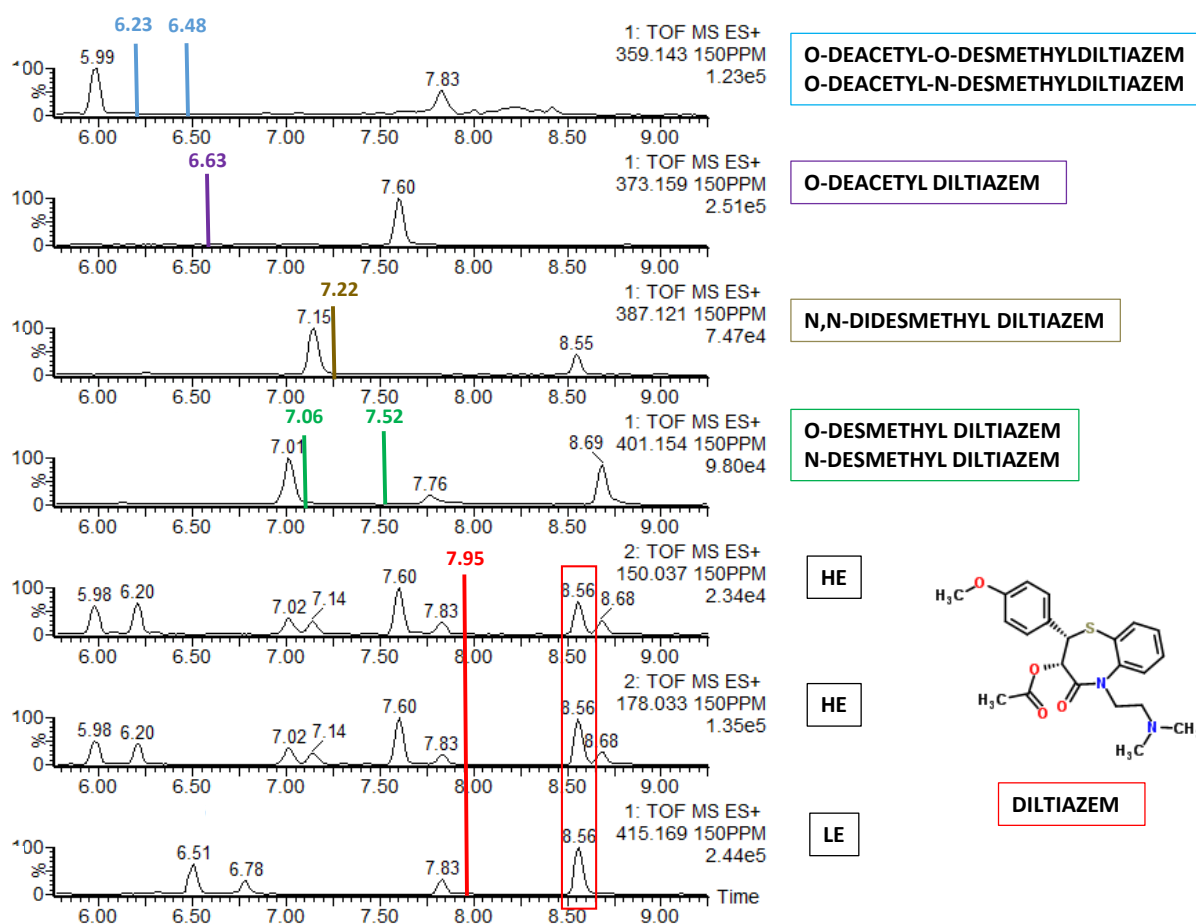


Figure 3.3: Tentative identification of diltiazem and six metabolites using common fragmentation pathways and t_R prediction. The solid lines are the predicted t_R while the circled peaks are those found based on common fragmentation with the colours relating to the colour of the borders of the compounds. Adapted from (Ibáñez et al., 2016).

In spite of the advances made using the ANN approach, it is impossible to have an absolutely precise means of prediction, not least due to reproducibility of liquid chromatography instruments, especially with respect to the stationary phase and the composition of the mobile phase as well as the matrix involved. Furthermore, as explained in **Scientific Article 6**, QSRR models are usually based solely on one laboratory's chromatographic instrument and its associated parameters, making it very difficult to transfer to other instruments and even slightly different parameters. As collaboration and information-sharing is very important in academia, it is therefore pertinent find a way to make retention time prediction models transferable and easy for anyone to use. Stopping short of insisting all laboratories use one instrument with all the same parameters, two possible solutions are:

- i) Indexing compounds, similar to the way Kovat's Index works in gas chromatography. This could be based on compounds not found in nature (i.e. deuterated standards) which cover the entire retention time range.
- ii) Performing retention time prediction on a range of columns and under different chromatographic conditions. This could then provide information on relevant descriptors needed for optimal prediction, which can be utilized by all laboratories.

CHAPTER 4

QUANTITATIVE DETERMINATION OF NEW PSYCHOACTIVE SUBSTANCES IN ENVIRONMENTAL WATER SAMPLES

4.1 Introduction

New psychoactive substances (NPS) are emerging narcotic or psychotropic drugs that are not controlled by the 1961 Single Convention on Narcotic Drugs or the 1971 Convention on Psychotropic Substances, but which may pose a public health threat. It must be noted that here the term ‘new’ does not necessarily refer to new inventions but to substances that have been recently become available in the market (UNODC, 2014). While the use of NPS appears to be limited overall, over the past ten years, the pace of emergence of new drugs has not diminished. Reports to the EU Early Warning System indicate that both the variety and quantity of NPS on Europe’s market are increasing. In 2014, more than 100 NPS were detected in Europe alone for the first time, and it is interesting to note how the new drugs on the market, mainly synthetic cannabinoids, stimulants, hallucinogens and opioids, mirror the established substances (i.e. cannabis, amphetamine, methamphetamine and cocaine) (EMCDDA, 2015). Initially not controlled under drug misuse legislation, NPS are easily available via the internet and “head shops”, described as ‘bath salts’ and ‘plant foods’ marketed as not for human consumption but which are indeed intended for recreational use.

The trouble with NPS is that they are continually changing due to the impact of legislation, with regulated compounds becoming “illicit drugs”, meaning possession and consumption is forbidden. New substances then fill this gap, and the process repeats. This means that determining which compounds to focus on for WBE approaches is very difficult, with there being a distinct possibility that the NPS in the methods are not found in any analysed samples. However, this may not be because the NPS are not being used, but rather the wrong biomarker is being searched. Little research has been made into the metabolism of NPS, and even when metabolism studies have been performed (Caspar et al., 2015; Chimalakonda et al., 2012; Mardal and Meyer, 2014; Moran et al., 2011), the standards (let alone deuterated standards) are sometimes difficult to obtain. This has led to WBE studies of NPS incorporating (predominantly) the parent compounds. (Borova et al., 2015; Kinyua et al., 2015a; Senta et al., 2015)

Between 2008-2013, the majority of NPS reported to the United Nations Office on Drugs and Crime were synthetic cannabinoids (28%), synthetic cathinones (26%) and phenethylamines (17%) (UNODC, 2014), on which NPS-based WBE approaches have been focusing.

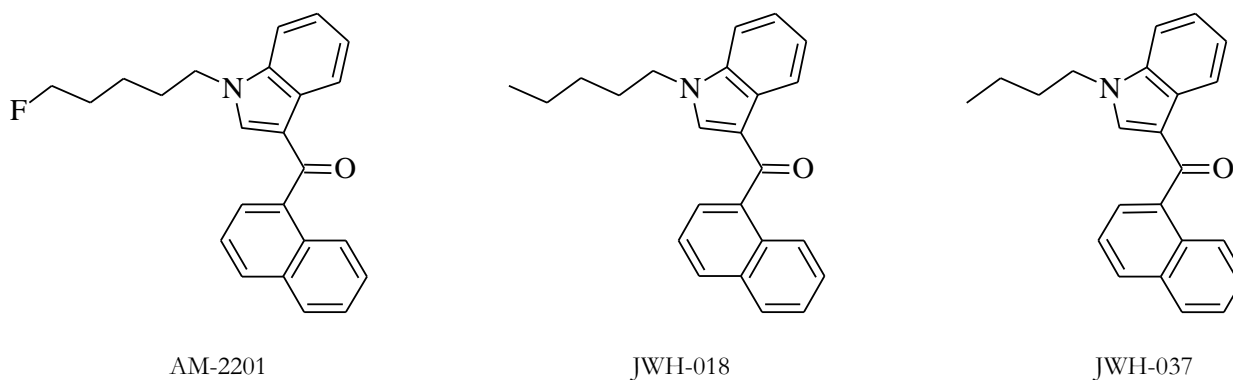


Figure 4.1: Structures of some synthetic cannabinoids

Synthetic cannabinoids are functionally but not necessarily structurally related to Δ^9 -tetrahydrocannabinol (THC), the active compound of cannabis, binding to the same cannabinoid receptors. They have become extremely popular as legal alternatives to cannabis and many have greater binding potency for CB1 and CB2 receptors than Δ^9 -THC and are full rather than partial agonists. (Hill and Thomas, 2016). They are sold as 'legal' smoking mixtures, containing many different compounds, some of which are non-psychoactive such as vitamin E, added in an attempt to 'mask' the active components. There has been much publicity around the synthetic cannabinoids, marketed as 'K2' and 'Spice' in the past ten years, with the European Monitoring Centre for Drugs and Drug Addiction convening an expert meeting on them in 2009 (EMCDDA, 2009). With legislation being introduced to halt the popularity of these mixtures, new analogues are being introduced as substances become restricted. Thus, the detection and identification of these compounds is very difficult and necessitate the use of HRMS to be able to differentiate the analogues (Grabenuer et al., 2012; Ibáñez et al., 2013). Nevertheless, for the most common synthetic cannabinoids (such as AM2201, JWH-018 and JWH-073, **Figure 4.1**), metabolism studies have been made (Chimalakonda et al., 2012; Moran et al., 2011), with the described metabolites used in quantitative WBE methods (Reid et al., 2014).

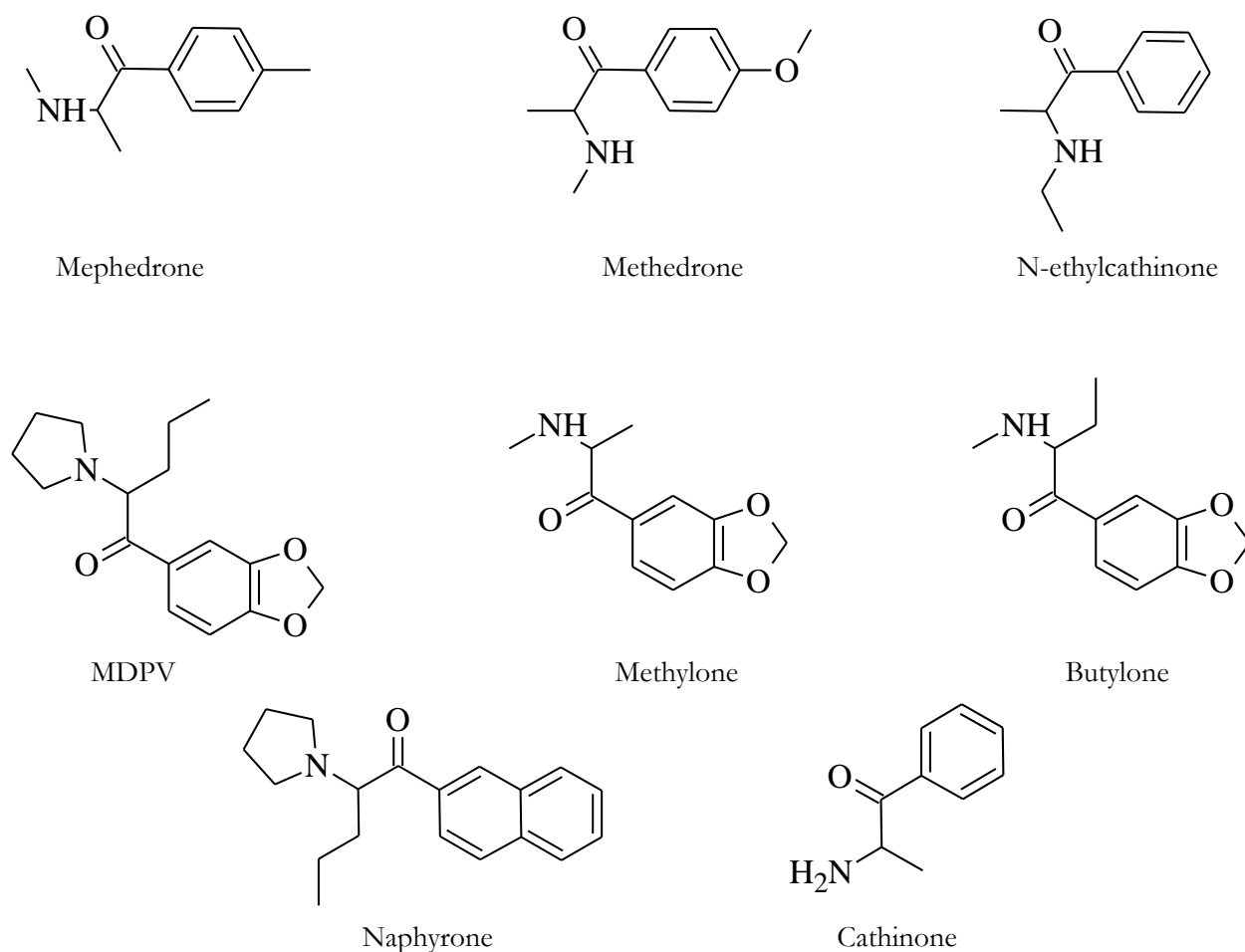


Figure 4.2: Structures of cathinone and the synthetic cathinone derivatives

Synthetic cathinones such as naphyrone, butylone, mephedrone, N-ethylcathinone, methedrone, methylone and methylenedioxypropylone (MDPV) are structurally related to cathinone (**Figure 4.2**), a natural compound found in the khat plant (*Catha edulis*), which is commonly chewed by people originating from countries around East Africa and the Arab Peninsula. They form a large family of compounds produced by substitutions at three locations of the cathinone molecule: phenyl ring, amino group and propanone terminus. Alkyl, alkoxy, alkylenedioxy, haloalkyl, cyclic groups and halides may be introduced to the molecule (Fornal, 2013). This has led to copious synthetic cathinones being developed and sold to the public in mixtures, with numerous synthetic cathinones currently identified and known (Ibáñez et al., 2014; Leffler et al., 2014; National Alliance for Model State Drug Laws, 2012; Zancajo et al., 2014). Due to the known toxicity of synthetic cathinones, legislation has been enacted in some countries, leading to a decrease in their use, however, in spite of these legal controls, some synthetic cathinones are still commonly abused (German et al., 2014; Hill and Thomas, 2016).

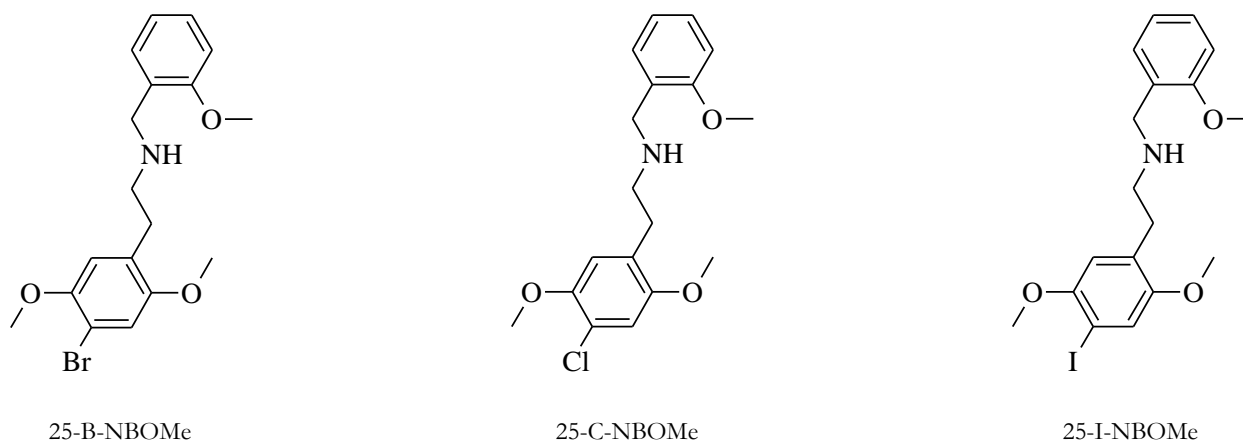


Figure 4.3: Structures of the “2C” phenethylamines

The N-2-methoxybenzyl phenethylamine (NBOMe) set of compounds are derived from a class of the well-known potent hallucinogenic **phenethylamines**, the so-called 2Cs (Maurer, 2010). Thus, they have a high potential for hallucinogenic effects with the risk of serotonergic toxicity. Among others, 2-(4-bromo-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine (25B-NBOMe), 2-(4-chloro-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine (25C-NBOMe), and 2-(4-iodo-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine (25I-NBOMe) (**Figure 4.3**) have been sold and consumed as so-called research chemicals. Among the NBOMe derivatives, 25I-NBOMe seems to be the most commonly used drug, seen as a ‘legal’ alternative to LSD. (Caspar et al., 2015; EMCDDA, 2015) These drugs may serve as replacements for the previously popular beta-keto derivatives of amphetamine – methcathinone and methedrone – as well as the methylenedioxy ring derivatives similar to MDMA (Ecstasy), methylone and methylenedioxypropylvalerone (MDPV), which are now controlled substances in Europe and the USA (Gibbons, 2012).

The metabolism of 25I-NBOMe has been studied (Caspar et al., 2015), with more than 60 metabolites found, including 26 in a human urine sample. With deuterated analogues of these metabolites very difficult to obtain, the parent compound is generally used for qualitative and quantitative analyses, and have been attributed to fatal intoxications (Kinyua et al., 2015b; Poklis et al., 2014; Tang et al., 2014).

The main difficulty associated with the quantitative analysis of new psychoactive substances (NPS) is the extremely low analyte concentration in the aquatic matrix. Within the context of WBE, traditional illicit drugs such as cocaine, MDMA, amphetamine and methamphetamine are found in the low ppb – ppt range, however, NPS are at even lower concentrations, due to their niche markets and ever-evolving compounds. This results in the need for very sensitive instrumentation, combined with specific sample pretreatment to minimise matrix effects. Furthermore, the complex nature of influent

wastewater requires filtration or centrifugation of samples prior to pretreatment. Filtration is generally made using glass fibre filters to remove any solid particulates. Due to the basicity of NPS, SPE is normally performed using cartridges with cationic exchange sorbents (e.g. Oasis MCX), with samples acidified (pH 2-3) prior to extraction. Comparisons have been made between the more generic Oasis HLB cartridges and the MCX, at differing pH, with MCX at pH 2-3 giving far better recoveries for the more basic synthetic cathinones and phenethylamines (Kinyua et al., 2015a; Senta et al., 2015; van Nuijs et al., 2013) while MCX has also been used for synthetic cannabinoids (Reid et al., 2014).

With the diverse range and ever-changing environment of NPS, one all-encompassing method has yet to be developed for their quantitative analysis in wastewater. In fact, a single LC technique has also yet to be decided upon, with both RP and HILIC used for their analysis. Within RP LC separation, various columns have been used, based on the NPS being analysed: from the generic C18 for the conventional illicit drugs and the NPS mephedrone, methylone and MDPV (Kankaanpää et al., 2014; van Nuijs et al., 2013) to the less hydrophobic C8 for the analysis of synthetic cannabinoids (Reid et al., 2014), pentafluorophenyl column for the analysis of synthetic cannabinoids, cathinones and amphetamines (Borova et al., 2015) and the polar ether-linked phenyl column for synthetic cathinones and amphetamines (Senta et al., 2015). Interestingly, HILIC has also been used for the analysis of the very polar phenethylamine-based compounds, which are retained well on HILIC columns (Kinyua et al., 2015a).

Nevertheless, the mobile phases utilised for all separations incorporate an aqueous and organic phase (typically methanol or acetonitrile). As mentioned previously, the addition of specific additives or buffers can enhance the chromatographic separation or influence the ionisation. The aqueous phase usually consists of acidified water (using 0.05-0.2% formic or acetic acid) and ammonium formate or ammonium acetate (5-10mM). The acidification improves the ionisation of the NPS, which are best ionised in positive ion mode.

As explained previously, sensitive determination of NPS is necessary for WBE, leading QqQ to be the instrument of choice, typically in SRM mode. This requires the preselection of all analytes, and all MS parameters are then optimised individually, such as cone voltage and collision energy. First of all, the ionisation mode and cone voltage are optimised, to define the precursor ion. Next, using the defined cone voltage and ionisation mode, variable collision energies are applied to determine product ions. At least two transitions (i.e. precursor ion > product ion) are necessary: one for quantification and one for confirmation. Typically the most abundant and sensitive transition is used for quantification. However, for oxygen-containing ephedrine-, cathinone- and phenethylamine-like NPS,

the most abundant transition corresponds mainly to the loss of water, which is a relatively non-specific transition. This can result in some matrix interferences, particularly in the complex matrix that is wastewater, which is full of organic hydroxyl-containing compounds. For such compounds, the second, less abundant, but specific transition should be used for quantification. In these cases, it is advisable to monitor additional transitions that can also be used for confirmation.

This chapter focusses on the use of UHPLC-QqQ-MS for the quantitative determination of NPS in influent wastewater.

4.2 Scientific Articles

This chapter includes one scientific article, “*Ultra-high performance liquid chromatography-tandem mass spectrometry determination of selected new psychoactive substances including synthetic cathinones and phenethylamines in influent wastewater*” (**Scientific Article 7**).

Scientific Article 7 presents a sensitive method for the quantification of ten NPS (N-ethylcathinone, methylone, butylone, MDPV, naphyrone, mephedrone, methedrone, 25-B-NBOMe, 25-I-NBOMe and 25-C-NBOMe) in influent wastewater. An investigation was made into the best extraction technique, with HLB and MCX SPE cartridges used, with MCX deemed to be superior for these basic compounds. Furthermore, the stability of these compounds was tested for up to 30 days at 20°C, 4°C and -20°C at both natural pH and pH 2. This method was used on wastewater samples from throughout Europe, with mephedrone, methylone and MDPV identified in several of the wastewater samples analysed.

These results have been written in the following article, submitted for publication:

Ultra-high performance liquid chromatography-tandem mass spectrometry determination of selected new psychoactive substances including synthetic cathinones and phenethylamines in influent wastewater of eight European cities R.Bade, L. Bijlsma, J. V. Sancho, J. A. Baz-Lomba, S. Castiglioni, E. Castrignanò, A.Causanilles, E. Gracia-Lor, B. Kasprzyk-Hordern, J. Kinyua, A.-K. McCall, A. L.N. van Nuijs, C.Ort, B. G. Plósz, P. Ramin, N. I. Rousis, Y. Ryu, K.V. Thomas, P. de Voogt, E. Zuccato, F. Hernández

4.2.1 Scientific Article 7

**ULTRA-HIGH PERFORMANCE LIQUID
CHROMATOGRAPHY-TANDEM MASS SPECTROMETRY
DETERMINATION OF SELECTED NEW
PSYCHOACTIVE SUBSTANCES INCLUDING
SYNTHETIC CATHINONES AND PHENETHYLAMINES
IN INFLUENT WASTEWATER OF EIGHT EUROPEAN
CITIES**

Richard Bade, Lubertus Bijlsma, Juan V. Sancho, J. A. Baz-Lomba, Sara Castiglioni, Erika Castrignanò, Ana Causanilles, Emma Gracia-Lor, Barbara Kasprzyk-Hordern, Juliet Kinyua, Ann-Kathrin McCall, Alexander L.N. van Nuijs, Christoph Ort, Benedek G. Plósz, Pedram Ramin, Nikolaos I. Rousis, Yeonsuk Ryu, Kevin V. Thomas, Pim de Voogt, Ettore Zuccato, Félix Hernández

Submitted for Publication

Ultra-high performance liquid chromatography-tandem mass spectrometry determination of selected new psychoactive substances including synthetic cathinones and phenethylamines in influent wastewater of eight European cities

Richard Bade^a, Lubertus Bijlsma^a, Juan V. Sancho^a, J. A. Baz-Lomba^{b,c}, Sara Castiglioni^d, Erika Castrignanò^e, Ana Causanilles^f, Emma Gracia-Lor^{a,d}, Barbara Kasprzyk-Hordern^e, Juliet Kinyua^h, Ann-Kathrin McCalli, Alexander L.N. van Nuijs^b, Christoph Ortⁱ, Benedek G. Plósz^j, Pedram Ramini, Nikolaos I. Rousis^d, Yeonsuk Ryu^{b,c}, Kevin V. Thomas^b, Pim de Voogt^{f,g}, Ettore Zuccato^d, Félix Hernández^{a,*}

a Research Institute for Pesticides and Water, University Jaume I, Avda. Sos Baynat s/n, E-12071 Castellón, Spain.

b Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway

c Faculty of Medicine, University of Oslo, PO box 1078 Blindern, 0316, Oslo, Norway

d IRCCS—Istituto di Ricerche Farmacologiche “Mario Negri”, Department of Environmental Health Sciences, Via La Masa 19, 20156 Milan, Italy

e University of Bath, Department of Chemistry, Faculty of Science, Bath BA2 7AY, UK

f KWR Watercycle Research Institute, Chemical Water Quality and Health, P.O. Box 1072, 3430 BB, Nieuwegein, The Netherlands

g Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, P.O. Box 94248, 1090 GE, Amsterdam, The Netherlands

h Toxicological Center, Department of Pharmaceutical Sciences, Campus Drie Eiken, University of Antwerp, Universiteitsplein 1, 2610 Antwerp, Belgium

i Eawag, Swiss Federal Institute of Aquatic Science and Technology, CH-8600, Dübendorf, Switzerland

j Department of Environmental Engineering, Technical University of Denmark, Miljøvej, Building 115, DK-2800 Kgs. Lyngby, Denmark

*Corresponding author: felix.hernandez@uji.es Tel.: +34 964 387366 Fax: +34 964 387368

Abstract

The use of new psychoactive substances (NPS) has grown in popularity over the past few years. Certain NPS are still commonly and even preferentially consumed even with the introduction of preventative legislation. With the objective to improve the knowledge on the use of NPS a rapid and very sensitive method was developed for the determination of ten priority NPS (N-ethylcathinone, methylenedioxypropylvalerone (MDPV), methylone, butylone, methedrone, mephedrone, naphyrone, 25-C-NBOMe, 25-I-NBOMe and 25-B-NBOMe) in influent wastewater. Sample clean-up and pre-concentration was made by off-line solid phase extraction (SPE) with Oasis MCX cartridges. Isotopically labelled internal standards were used to correct for matrix effects and potential SPE losses. Following chromatographic separation on a C18 column within 6 minutes, the compounds were measured by tandem mass spectrometry (triple quadrupole) in positive ionisation mode. The method was optimised and validated for all compounds, with limits of quantification being 1 or 5 ng L⁻¹, with the method having been fully validated in samples spiked at these levels. An investigation into the stability of these compounds in influent wastewater was also performed, showing that, following acidification at pH 2, all compounds were relatively stable for up to 7 days. The method was then applied to influent wastewater samples from eight European countries, in which mephedrone, methylone and MDPV were detected. This work reveals that although NPS use is not as extensive as for classic illicit drugs such as cocaine, cannabis, amphetamines, the application of a highly sensitive analytical procedure makes their detection in wastewater possible. The developed analytical methodology forms the basis of a subsequent model-based back-calculation of abuse rate in urban areas (i.e. wastewater-based epidemiology) in the future.

Key Words: New psychoactive substances, Ultra high-performance liquid chromatography, Triple Quadrupole, Wastewater, Stability, Matrix Effects

1. Introduction

New psychoactive substances (NPS) are emerging narcotic or psychotropic drugs that are not controlled by legislation, but which may pose a public health threat. It must be noted that in this context, the term ‘new’ does not necessarily refer to new inventions but to substances that have been recently become available [1]. The use of NPS has grown rapidly over the past decade and there have been increasing reports of the availability and manufacture of such substances, with the number of NPS reported globally more than doubling between 2009–2013 (UNODC 2014). In 2014 alone, 101 NPS were for the first time reported to the EU Early Warning System [2]. Given the nature of the NPS market, with developers limited only by their imagination and ability to sidestep legislation [3] there is a sustained need to analyse the extent of the NPS problem.

The analysis of wastewater to estimate (illicit) drug consumption based on biomarkers, has traditionally focussed mainly on the most commonly used illicit drugs - cocaine, cannabis, amphetamine, methamphetamine, 3,4-methylenedioxymethamphetamine (ecstasy, MDMA) [4,5], leaving a large information gap on other illicit drugs and NPS. Little research has been made on NPS and their suitable biomarkers, let alone their stability. In the few papers that have been published until now on the determination of NPS in wastewater, the target analytes included are commonly the synthetic cathinones mephedrone, methylone and MDPV [6–15]. Within these studies, the most commonly detected NPS in wastewater are mephedrone and MDPV, generally found at the low ng/L range.

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is commonly the technique of choice for the quantitative determination of illicit drugs in wastewater, due to the low concentrations involved and the high sensitivity of the instrument. When using this technique, several analytical aspects are of relevance in order to report accurate data. In addition to the required validation at realistic concentrations that can be found in the samples, relevant issues must be considered, such as the way to correct/minimize matrix effects, and the proper identification of the compound detected. The use of isotopically-labelled internal standards (ILIS) is one of the most efficient tools to correct for matrix effects as well as potential losses from solid phase extraction (SPE). The confirmation of identity of the compounds is essential in order to avoid reporting false positives. By utilising LC-MS/MS instruments in selected reaction monitoring (SRM) mode, at least two transitions should be incorporated in the method (one for quantification and the other(s) for confirmation). However, the specificity of the transitions should be taken into account, as non-specific transitions (such as the loss of water) can lead to false negatives due to the non-compliance of ion ratios [16].

The purpose of this study was to develop and validate a sensitive LC-MS/MS method for the quantitative determination of a number of NPS of the synthetic cathinone and phenethylamine families: butylone, ethylone, methylone, naphyrone, methedrone, methylenedioxypropylone (MDPV), mephedrone, 25-I-NBOMe, 25-

C-NBOMe and 25-B-NBOMe. These compounds were selected on the basis of their frequent detection in analytical, forensic and toxicological studies [6–10,12,13,17–19] as well as reports from the EMCDDA [20] and UNODC [1]. The developed method, using Oasis MCX SPE cartridges for sample pre-treatment, followed by UHPLC-MS/MS measurement, has been applied to influent wastewater samples from around Europe, with an additional study on stability being made. Special emphasis is placed on the reliable confirmation of the NPS detected in water, with up to three SRM transitions being acquired, which, together with ion ratios, allowed simultaneous detection, quantification and confirmation of positive samples.

2. Experimental

2.1 Chemicals and Materials

Reference standards of all analytes: N-ethylcathinone, MDPV, methylone, butylone, methedrone, mephedrone, naphyrone, 25-C-NBOMe, 25-I-NBOMe and 25-B-NBOMe were purchased from Cerilliant (Round Rock, TX, USA) at a concentration of 1 mg/mL in methanol for the analytes. Standard stock solutions of each compound were prepared at 100 mg/L.

Mixed working solutions containing all analytes were prepared from stock solutions at different concentrations by appropriate dilution with water, and were used for preparation of the aqueous calibration standards and for spiking samples in the validation study.

Deuterated analogues: MDPV-d8, methylone-d3, butylone-d3, mephedrone-d3, naphyrone-d5, 25-C-NBOMe-d3, 25-I-NBOMe-d3 and 25-B-NBOMe-d3 were also purchased from Cerilliant (Round Rock, TX, USA) at a concentration of 100 µg/mL in methanol. Standard stock solutions of each compound were prepared at 10 mg/L. Mixed standard solutions containing all compounds were made with appropriate dilution with water.

All standard solutions were stored in amber glass bottles in the dark at -20°C.

HPLC-grade methanol (MeOH), ammonium acetate, ammonia solution (25 %) and formic acid (HCOOH, 98–100 %) were acquired from Scharlau (Barcelona, Spain) HPLC-grade water was obtained by purifying demineralised water in a Milli-Q plus system from Millipore (Bedford, MA, USA). SPE cartridges used were Oasis HLB 3cc (60mg) and Oasis MCX 6cc (150 mg) from Waters (Milford, MA, USA)

2.2 Samples

A number of different influent wastewater (IWW) samples were utilised in the development and validation of the present method: from Zurich, Switzerland; Copenhagen, Denmark and Castellon, Spain. The developed method was applied to IWW samples. The 24-h composite samples were taken in March 2015 from Zurich, Switzerland; Copenhagen, Denmark; Oslo, Norway; Castellon, Spain; Milan, Italy; Brussels, Belgium and

Utrecht, The Netherlands and Bristol, United Kingdom. All samples were collected in high density polyethylene bottles, transported to Castellon and stored in the dark at -20°C until pre-treatment.

2.3 Instrumentation

A Waters Acquity UHPLC system (Milford, MA, USA) was interfaced to a triple quadrupole mass spectrometer (Xevo TQS, Waters Micromass, Manchester, UK) equipped with Z-Wave devices and an electrospray ionization interface (ESI) operated in positive-ion mode. The chromatographic separation was performed using an Acquity UPLC BEH C₁₈ column, 1.7 μm, 50 mm × 2.1 mm (i.d.) (Waters) at a flow rate of 0.3 mL min⁻¹. The mobile phases used were water with 5 mM ammonium acetate and 0.01% formic acid (solvent A) and MeOH with 0.01% formic acid (solvent B). The percentage of B changed linearly as follows: 0 min, 10 %; 3 min, 90 %; 3.5 min, 90 %; 3.6 min, 10 %; 6 min 10 %, equilibration of the column. Cone gas as well as desolvation gas was dry nitrogen, with flows set to 150 and 800 L h⁻¹, respectively. For operation in the MS/MS mode, collision gas was argon 99.995 % (Praxair, Madrid, Spain) with a pressure of 4×10⁻³ mbar in the collision cell (0.15 mL min⁻¹). Other parameters optimized were: capillary voltage, 3.0 kV; source temperature, 150 °C and desolvation temperature, 650 °C. Dwell times of 0.01 s/transition were automatically selected.

All data were acquired and processed using MassLynx v4.1 software (Waters, Manchester, UK).

2.4 Sample Pre-treatment

All water samples were acidified to pH 2 with formic acid (98%), then centrifuged for 5 minutes at 6000 rpm. SPE cartridges were conditioned with methanol (6mL), water (3mL) and acidified water (pH 2 with formic acid, 3mL). The IWW samples (100mL) were spiked with mixed surrogate ILIS to give a final in sample concentration of 20 ng L⁻¹, then percolated through the cartridges at ±5 mL min⁻¹. The cartridges were then washed with 5 mL acidified methanol (pH 2 with formic acid) and dried under vacuum for 10 minutes. The analytes were eluted with 5 mL MeOH (2% ammonia), evaporated to dryness at 35°C under nitrogen and reconstituted to 1mL with 10:90 methanol:water.

Analyses were performed by injecting 3 μL of the final extract in the UHPLC-MS/MS system.

2.5 Quantification and Method Validation

SRM mode was used to acquire all data, with the precursor ion of each compound being the protonated molecule. In general, the most abundant product ion of each target analyte was used for quantification, with one or two additional product ions and their ion ratios used for confirmation. In order to confirm the compounds detected, LC retention time was compared to that of the reference standards (within ± 0.1 minutes). These steps are in line with the SANCO guidelines for analytical quality control and validation procedures [21], which has been utilised previously by research groups for quantification of illicit drugs in wastewater [22]. Each

compound was quantified using its corresponding ILIS as surrogate internal standard, except N-ethylcathinone and methedrone, which were quantified using butylone-d3 and methylone-d3, respectively. These ILIS were chosen based on their ability to correct for matrix effects as well as having similar retention times and chemical structures closely related to the analyte of interest.

The performance of the method was evaluated in terms of linearity, limits of detection and quantification, accuracy and precision.

The linearity was studied by analysing standard solutions in solvent in triplicate at seven concentration levels ranging from 0.5-100 ng L⁻¹. Satisfactory linearity using weighted (1/X) least squares regression was assumed when the correlation coefficient (r) was >0.99.

For limits of quantification and detection (LOQ and LOD), two different concentration levels (1 and 5 ng L⁻¹) were tested so as to provide a more accurate measure. The limit of quantification (LOQ) was established as the lowest level for which the method was fully validated using spiked IWW samples with satisfactory accuracy based on recovery experiments (between 70-120%) and precision (relative standard deviation (RSD) ≤20%). Furthermore, at least two MS/MS transitions were required at the LOQ level for confirmation. All recovery experiments were performed in quintuplicate.

The limit of detection (LOD) was estimated using blank samples spiked at the lowest analyte concentration tested, based on a signal-to-noise level of three from the quantification SRM chromatogram.

2.6 Stability experiments

Stability experiments were performed under different conditions (IWW at natural pH and acidified to pH 2, filtered influent wastewater at natural pH and acidified to pH 2 and MilliQ water) and at three different temperatures (20°C, 4°C and -20°C) for 24 hours, 48 hours, 7 days, and 30 days (at -20°C only). The samples were each initially spiked at a concentration level of 100ng L⁻¹. One mL of each sample was collected at the time intervals outlined above and spiked with ILIS (100 ng L⁻¹) just before analysis.

3. Results and Discussion

3.1 Selection of Compounds

Only parent compounds were selected for this study. This is based on the previous metabolism studies performed on the selected ten as well as related NPS in human and rat urine [18,23–27], which showed that in spite of the number of metabolites identified, the parent compound could still be detected in urine, thereby giving credence to their inclusion as a biomarker in this method. Furthermore, it has been noted that the metabolic patterns of selected phenethylamine-based designer drugs have indicated rather slow metabolism rates, with parent drugs being the key biomarkers of consumption [8,28], somewhat expected due to their structural similarity to amphetamine and MDMA [5].

In addition, for the phenethylamine-based 25-X-NBOMe compounds, the parent compound is generally used in both qualitative and quantitative toxicological studies [29–31].

3.2 Instrument Optimisation

Initially, all cone voltages were optimised concurrently for all compounds, using a mixed standard solution (10 $\mu\text{g L}^{-1}$), with the cone voltage ranging from 10-50 V at 10 V intervals. For all compounds, the precursor ion was $[\text{M}+\text{H}]^+$. Collision energies were optimised for each compound independently using a collision energy ranging from 10-50eV at 10eV intervals. The most sensitive transition was typically selected for quantification (Q) while an additional one (or two) were selected for confirmation (q). Some compounds only had two product ions of significant intensity, meaning that only two transitions (i.e. Q and q1 only) could be monitored (**Table 1**). Non-specific transitions such as the loss of water were avoided as much as possible (except for methylene, mephedrone and N-ethylcathinone where they were used as confirmation transitions) in order to reduce the likelihood of false positives [22]. For the three 25-X-NBOMe compounds, the same product ions were seen (m/z 121 and 91), corresponding to the cleavage of the methoxybenzyl moiety (m/z 121) and the further loss of the methoxy group producing the tropylium ion (m/z 91). These transitions are not very specific, and can come from any compound with a methoxybenzoyl moiety, however as there were no other transitions available for these compounds, they had to be used as quantification and confirmation transitions. Regarding the ILIS, only a single transition was monitored.

The chromatographic separation was initially based on a mobile phase composed of water:MeOH, which resulted in very poor peak shapes. Different concentrations of ammonium acetate buffer and pH (formic acid) were investigated. The effects of pH and mobile phase ionic strength on the peak shapes, resolution and efficiencies were evaluated by varying the buffer concentration. An optimal mobile phase consisting of water with 5 mM ammonium acetate and 0.01% formic acid (solvent A) and MeOH with 0.01% formic acid (solvent B) was finally selected.

Table 1: MS/MS optimised conditions for all studied compounds

Compound	Retention time (mins)	Precursor ion [M+H] ⁺ (m/z)	Cone Voltage (V)	Collision Energy (eV)	Product Ions (m/z)	LOQ ^a (ng L ⁻¹)	LOD ^b (pg L ⁻¹)	q/Q ratio (RSD %)
25-B-NBOMe	3.76	380.1	40	20	121.0	1	100	
				30	91.0			0.4 (6)
25-C-NBOMe	3.70	336.1	20	20	121.1	5	70	
				40	91.0			0.6 (6)
25-I-NBOMe	3.86	428.1	20	20	121.0	5	100	
				50	91.0			0.5 (6)
Butylone	2.68	222.1	20	20	174.0	5	200	
				20	146.0			0.3 (6)
				30	131.2			0.3 (7)
Mephedrone	2.77	178.1	30	20	145.2	5	200	
				20	119.0			0.1 (11)
				10	160.0			1.3 (7)
Methedrone	2.57	194.1	20	20	161.1	5	200	
				20	146.0			0.4 (4)
MDPV	3.11	276.2	50	20	175.1	1	200	
				20	126.0			0.6 (5)

				20	205.0			0.8 (9)
Methylone	2.36	208.1	50	20	160.0	5	50	
				20	132.1			0.4 (6)
				10	190.0			0.5 (11)
Naphyrone	3.71	282.2	50	20	141.2	1	50	
				20	211.1			0.5 (6)
				40	127.4			0.2 (6)
N-ethylcathinone	2.42	178.1	20	10	130.0	5	200	
				20	117.1			1.1 (7)
				30	160.1			0.7 (10)
25-B-NBOMe-d ₃	3.76	383.3	40	20	124.0			
25-C-NBOMe-d ₃	3.70	339.1	20	20	124.0			
25-I-NBOMe-d ₃	3.86	431.1	20	20	124.0			
Butylone-d ₃	2.68	225.1	20	20	177.1			
Mephedrone-d ₃	2.77	181.1	30	20	145.2			
MDPV-d ₈	3.11	284.2	50	20	205.0			
Methylone-d ₃	2.36	211.1	50	20	163.1			
Naphyrone-d ₅	3.71	287.2	50	20	141.2			

^aLOQ objective. Lowest concentration tested for the method being fully validated with satisfactory results.

^bLOD estimated for a signal-to-noise level of three from the quantification SRM chromatogram of blank samples spiked at the lowest analyte concentration tested.

3.3 Optimisation of extraction procedure

An investigation into the filtration losses of these compounds was made in samples at both natural and pH 2, using a vacuum filter (Whatman, 0.45 μ m mixed cellulose ester membrane). It was found that for naphyrone and the 25-X-NBOMe compounds, less than 5% of the compounds were recovered following filtration, while for MDPV a recovery of 55-71% was found. (**Table S1**) With these results, filtration was not employed in this method, instead we only performed centrifugation.

Two cartridges were evaluated in the optimisation of the SPE procedure: Oasis HLB and Oasis MCX, which are two of the most popular cartridges for the extraction of different types of illicit drugs in environmental waters [32]. The backbone of the Oasis sorbents consists of apolar moieties (benzyl groups, aliphatic chains) and polar groups (pyrrolidone). HLB is only synthesised from these two monomers and neutral. MCX has extra sulfonic acid functional groups (1.01mmol g⁻¹) implemented at a fixed location and contain permanently charged groups at any relevant pH. For both cartridges, 100 mL of IWW was spiked with the compounds (100 ng L⁻¹) before and after extraction and the recovery was used to determine the better cartridge for further optimisation. No pH adjustment was made for the Oasis HLB cartridges, while for the MCX, samples were acidified (pH 2) in order to protonate all basic analytes, thereby enhancing their ionic interactions with the sorbent [8]. This led to a reduction in matrix interferences, giving higher sensitivity and peak shapes for the majority of the compounds (**Figure 1**). The SPE procedure followed for the HLB cartridges was the same as in previously published studies. [33,34]

Higher recoveries were obtained when using Oasis MCX cartridges (**Figure 2**) and therefore these were selected for further optimisation. For the washing step, water, pure MeOH and acidified MeOH (pH 2) were tested. Drying under vacuum (or not) was also tested, to determine the optimal washing/drying steps (**Figure 2**).

As seen in **Figure 2**, the optimal SPE procedure involved washing with 5mL acidified methanol and drying for 10 minutes, with recoveries ranging from 83%-93% (**Figure 2**, MCX 3).

In previous methods for NPS using MCX cartridges, different percentages of ammonia in methanol have been used for elution, ranging from 0.5% to 4% [7,8,11,12]. A compromise of 2% was utilised in this work, and showed good recoveries for all compounds. An elution volume of 5mL of 2% ammonia in MeOH solution was finally chosen. This was based on the comparison of the recoveries of 5mL and 10mL elution volumes, with recoveries only 2% less for the 5mL than the 10mL.

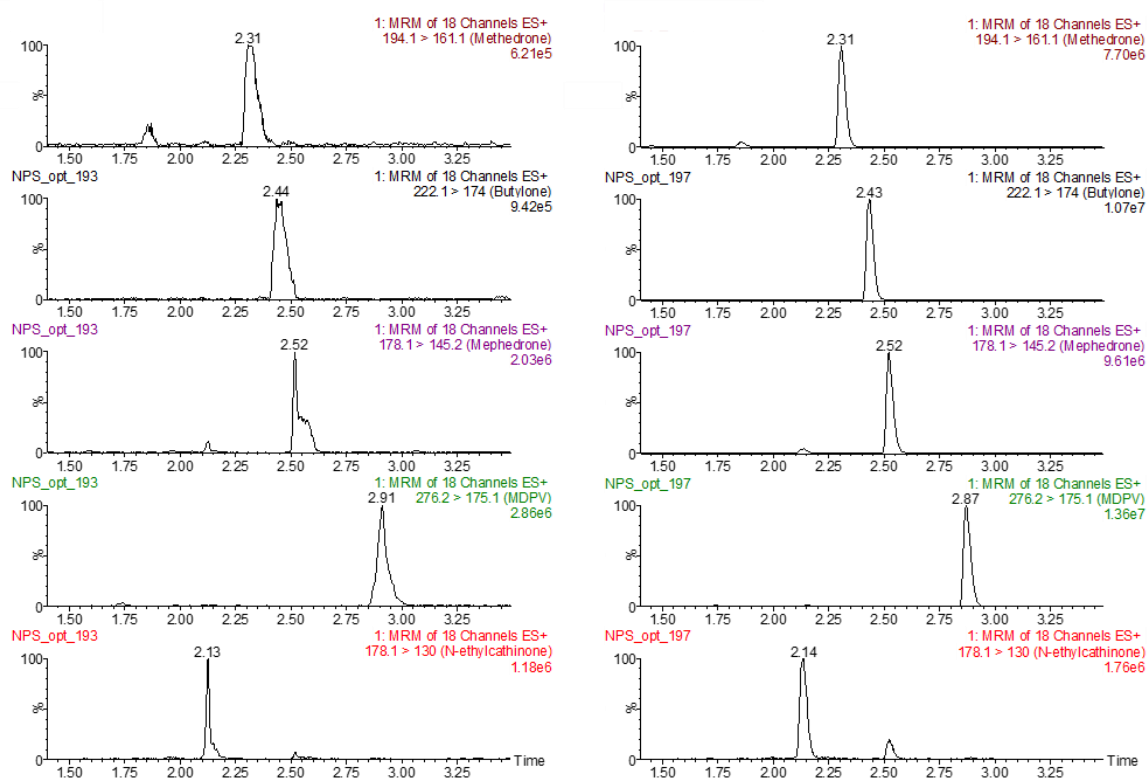


Figure 4: SRM of a selection of compounds, after extraction with HLB (LEFT) and MCX (RIGHT) cartridges. From top to bottom: methedrone, butylone, mephedrone, MDPV and N-ethylcatinone.

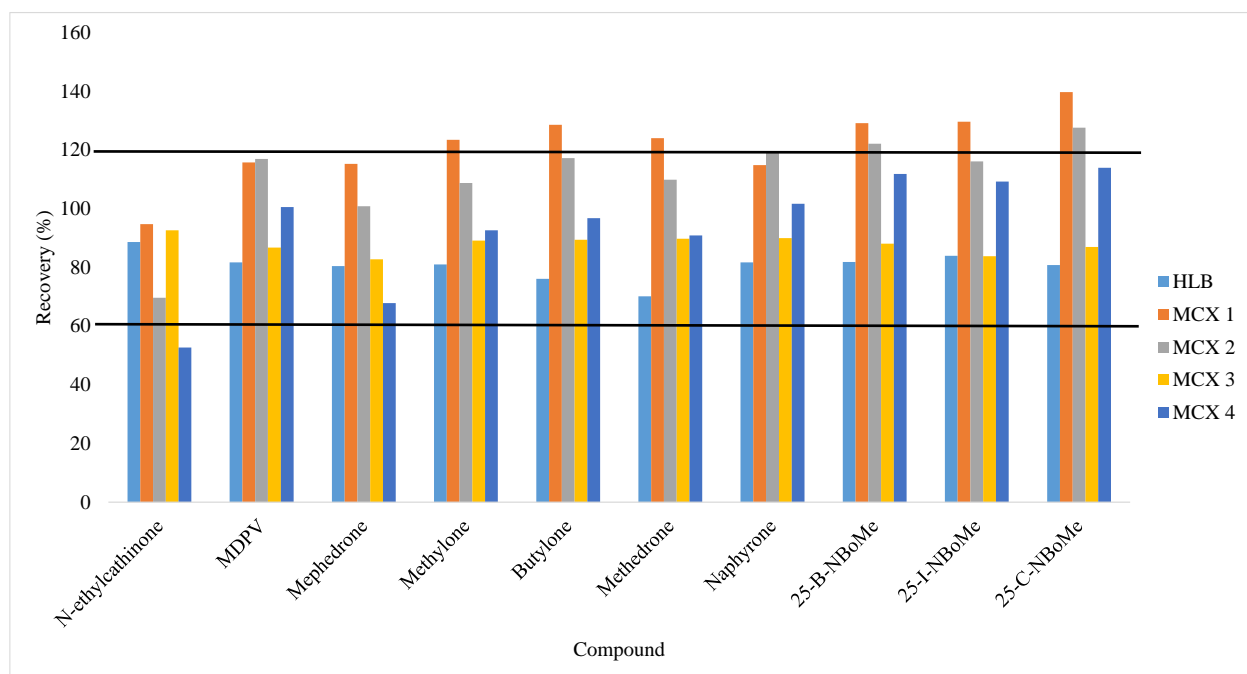


Figure 5: Recoveries for all compounds using HLB and MCX cartridges. Legend shows different washing/drying steps used. Black lines are shown at the threshold limits for satisfactory recovery (70-120%)

3.4 Method Validation

Before applying the developed methodology to IWW samples, the method was fully validated for linearity, precision and accuracy (recovery), LOD, LOQ and q/Q ratio (**Tables 1 and 2**). A calibration curve of seven points was made in solvent from 0.5-100 ng L⁻¹. Correlation coefficients were greater than 0.99 for all compounds, except for naphyrone and the NBOMe compounds, which had values between 0.97-0.98. For precision and accuracy, blank IWW samples were spiked at four concentration levels (1 ng L⁻¹, 5 ng L⁻¹, 10 ng L⁻¹ and 50 ng L⁻¹), and analysed in quintuplicate. These two sets corresponds to the different LOQs of some of the compounds (either 1 ng L⁻¹ or 5 ng L⁻¹, see **Table 1**). Accuracy was measured by means of recovery, with all compounds showing satisfactory recovery (between 70-120%), and precision (all within RSD 20%).

LOQs were estimated from a non-statistical point of view, utilising an LOQ objective: the lowest level in sample for which the method was fully validated in terms of accuracy and precision. This criterion has been utilised in the past in published results [22] and is used in other fields, such as pesticide residue analysis [21]. This gives a more realistic LOQ value, albeit generally higher than those estimated from statistical criteria (i.e. signal-to-noise of 10 at the lowest validation level tested).

Average intensity q/Q ratios were calculated from reference standards in solvent of different concentrations (**Table 1**). With RSD% values $\leq 11\%$ for all q/Q ratios, there was high confidence that the ratios are not concentration dependent. These were compared to those experimentally obtained from sample extracts spiked at the lowest level validated. The aim of this was to test the robustness of these values and to check for potential matrix interferences that could affect the q/Q ratios and consequently, the confirmation process. As **Table 2** shows, average q/Q deviations were all below 30% [21], except for 25-I-NBOMe and 25-C-NBOMe at the lowest level of 1 ng/L. This finding returns to the previous discussion of the non-specificity of the NBOMe transitions, whereby although these transitions could be seen at this level, matrix interferences let the q/Q deviations far exceed the threshold of the SANCO guidelines, which were being followed in this work. This led to the LOQ for these two compounds being increased to 5 ng L⁻¹.

Matrix effects are a known drawback of LC-MS/MS, which can adversely affect the compounds of interest, leading to signal enhancement or suppression, thereby affecting the quantification process. These effects are most pronounced in more complex matrices, such as IWW. In this work, matrix effects were calculated by comparing the peak areas of Set 1 (SPE eluates (IWW previously extracted) spiked with the mixed standard and ILIS solutions) with Set 2 (mixed standard solution in solvent, including ILIS). All samples were performed in triplicate and averaged to perform the following calculations [35]:

$$\text{Matrix effects (\%)} = \frac{\text{average peak area (Set 1)}}{\text{average peak area (Set 2)}} \times 100$$

These calculations showed the true impact of the matrix on the compounds, with all showing significant matrix enhancement (**Table 2**). To correct for these interferences, ILIS were included in the calculation:

$$\text{Matrix effects (\%)} = \frac{\text{average peak area (Set 1,ILIS)} / \text{average peak area (Set 2,ILIS)}}{\text{average peak area (Set 1)} / \text{average peak area (Set 2)}} \times 100$$

Including ILIS to the calculations led to corrected matrix effects under 10% for all compounds.

Additional data pertaining to the use of ILIS is shown in **Table 2**, for SPE recovery and matrix effects, showing how well the ILIS are able to correct for matrix interferences.

Table 2: Method validation for all compounds in IWW at 2 validation levels (all values are in %), together with recovery and matrix effects before and after correction with internal standard (n=5). Linearity is also shown.

Compound	Matrix effects (%) ^a		q1/Q ratio		q2/Q ratio		SPE Recovery (%)				Linearity
			Low ^b	High ^c	Low ^b	High ^c	Low ^b		High ^c		
		No ILIS						No ILIS		No ILIS	
25-B-NBOMe	-4	60	-11 (12)	-9 (4)	-	-	121 (7)	87 (13)	83 (4)	57 (16)	0.9722
25-C-NBOMe	-7	50	-7 (8)	-2 (0)	-	-	86 (9)	51 (14)	89 (7)	73 (15)	0.9737
25-I-NBOMe	-6	66	5 (8)	-13 (1)	-	-	111 (9)	85 (16)	77 (6)	78 (16)	0.9722
Butylone	-6	29	-13 (12)	-4 (6)	24 (26)	-11 (4)	79 (10)	49 (13)	83 (4)	55 (8)	0.9923
MDPV	-7	48	6 (8)	10 (4)	17 (13)	9 (8)	122 (10)	94 (8)	74 (9)	52 (10)	0.9900
Mephedrone	-9	26	-6 (11)	12 (4)	0 (5) ^d	15 (4) ^d	71 (10)	40 (8)	94 (7)	52 (6)	0.9914
Methedrone	-10	31	-12 (4)	-7 (6)	-	-	73 (8)	49 (7)	80 (5)	53 (7)	0.9923
Methylone	-6	30	6 (7)	-1 (8)	^d	^d	71 (8)	30 (4)	89 (4)	36 (7)	0.9919
Naphyrone	-10	62	-5 (5)	-7 (9)	-5 (9)	-13 (3)	111 (10)	75 (10)	76 (8)	50 (15)	0.9883
N-ethylcathinone	3	27	-33 (21)	1 (7)	^d	5 (6) ^d	113 (7)	47 (12)	84 (7)	34 (5)	0.9975

ILIS (Isotopically-labelled Internal Standard)

^a: for matrix effects, a negative value (-) denotes matrix suppression, while positive values indicate matrix enhancement.

^b: Low refers to the LOQ value of all compounds: 1 ng L⁻¹ for MDPV, Naphyrone and 25-B-NBOMe and 5 ng L⁻¹ for all other compounds

^c: High refers to the 10xLOQ value of all compounds: 10 ng L⁻¹ for MDPV, Naphyrone and 25-B-NBOMe and 50 ng L⁻¹ for all other compounds

^d: These transitions were for a loss of water. The non-specificity associated with these transitions combined with the complex IWW matrix led to some of these transitions being non-observed/interfered. However, as they were all q2 transitions, the method could still be validated based on q1.

3.5 In-sample Stability

Stability of analytes is a key matter in any analysis, particularly when dealing with very low analyte concentrations in complex matrices such as wastewater. It is imperative that target analytes do not degrade during the 24-hour composite samples collection or during their storage before extraction. Composite samples are normally collected at 4 °C and are stored either at that temperature, or at -20 °C, depending on the time lapsing between collection and extraction[8], while extraction is usually carried out at room temperature. For these reasons, the stability of the target analytes in this study was studied at 4 °C, -20 °C and 20 °C. The study was carried out in both raw and filtered wastewater (as some research groups filter their samples prior to storage) as well as in Milli-Q water. Furthermore, the impact of acidification was also examined, with the samples examined at both natural pH and at pH 2.

In this section, low (60-100% transformation), medium (20-60% transformation) and high stability (0-20% transformation) is defined as by McCall et al [36], with all results being normalised to $t=0$. It must be stated that there was little difference between the filtered and unfiltered wastewater for most analytes, with the compounds less stable at natural pH. At 20 °C, N-ethylcathinone and mephedrone, showed low stability after 24 hours. However, at pH 2, all compounds exhibited high stability at all temperatures, for at least 7 days.

The results show that if the samples are not acidified they should be refrigerated (at 4 °C), or frozen (at -20 °C), and analysed within 7 days to ensure no significant degradation has occurred. Alternatively, if samples are acidified to pH 2, all investigated compounds are more stable and can even be maintained at room temperature for up to 7 days. However, after 30 days, there were some distinct stability issues with some compounds: mephedrone, methedrone and butylone all exhibiting low stability (up to 84% transformation); N-ethylcathinone, MDPV and methylone showing medium stability while the 25-X-NBOMe compounds and naphyrone still had high stability (less than 10% transformation) (**Figure 3**). All other results of the stability tests (i.e. in filtered wastewater and Milli-Q water) can be found in the supporting information (Figures S1-S3). These results also mirror those of the recent studies of Senta et al [8], and Reid et al [12], who investigated the stability of some of these compounds.

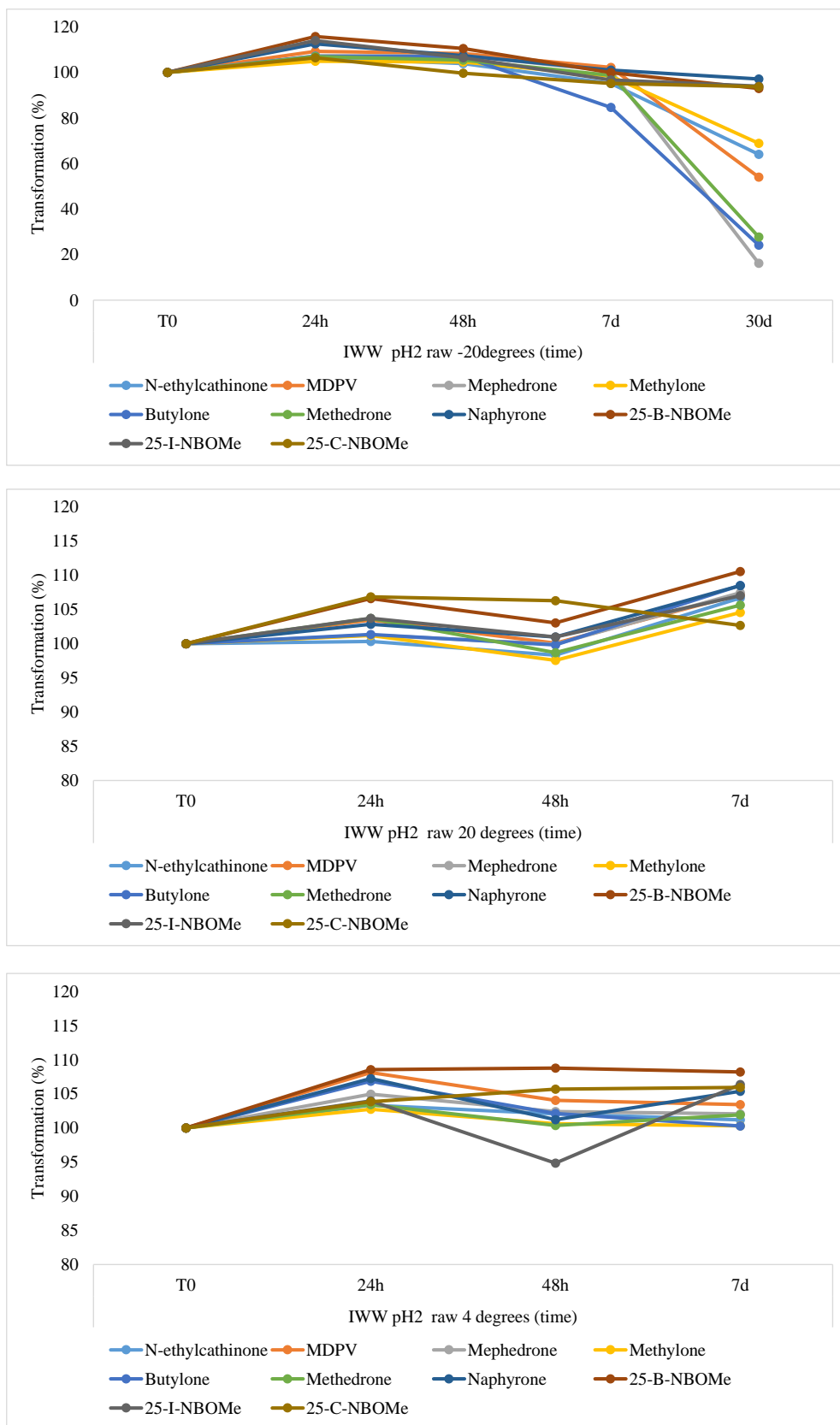


Figure 3: Stability of all compounds in raw IWW at (pH 2) at 20 °C (TOP), 4 °C (MIDDLE) and -20 °C (BOTTOM). Transformation (y axis) shows the value of the analyte remaining (i.e. a value of 20% means a transformation of 80%)

3.6 Application to Real Samples

A selection of samples from throughout Europe were investigated for the presence of NPS. In total, IWW from eight cities were investigated: Bristol, United Kingdom; Brussels, Belgium; Castellon, Spain Oslo, Norway; Copenhagen, Denmark; Milan, Italy; Utrecht, The Netherlands; Zurich, Switzerland, and samples were from weekends (Saturday and Sunday) and weekdays (Monday-Friday).

In every sequence of analysis, water samples were injected between two calibration curves. Two quality control samples (QCs), i.e. a blank water sample (previously analyzed) spiked at LOQ and 10×LOQ levels, were also analysed. QC recoveries were considered satisfactory when they were in the range of 70–120% for each analyte.

Table 3: Concentrations (in ng L⁻¹) of investigated NPS in IWW throughout Europe (weekday/weekend)

Compound	South West UK	Brussels	Castellon	Copenhagen	Milan	Oslo	Utrecht	Zurich
N-ethylcathinone	-	-	-	-	-	-	-	-
MDPV	-	-	-	-	-/3	-/2	-	6/-
Mephedrone	61/106	d/-	-	5/5	-	d/d	-/d	-/d
Methylone	-/d	-	-	-/12	-	-	7/5	-
Butylone	-	-	-	-	-	-	-	-
Methedrone	-	-	-	-	-	-	-	-
Naphyrone	-	-	-	-	-	-	-	-
25-B-NBOMe	-	-	-	-	-	-	-	-
25-I-NBOMe	-	-	-	-	-	-	-	-
25-C-NBOMe	-	-	-	-	-	-	-	-

d: detected, at a concentration below LOQ; -: not detected (<LOD)

As shown in **Table 3**, mephedrone was the most commonly identified NPS, with concentrations ranging from <5 ng L⁻¹ (Brussels, Oslo, Copenhagen and Utrecht) to 106 ng L⁻¹ (South West UK). The high concentrations found in Bristol are similar to another study in the United Kingdom, which found concentrations up to 114 (± 15) ng L⁻¹. [15] These data also follow that from the UNODC, stating that mephedrone has the highest prevalence rate of any NPS in the United Kingdom, and behind only cannabis, cocaine and ecstasy in terms of drug use [1]. The only other compounds confirmed in our study were methylone and MDPV, with concentrations ranging from below LOQ to 12 ng L⁻¹ (methylone) and 6 ng L⁻¹ (MDPV), with these low concentration in line with previous studies in Croatia and Switzerland [7,8]. These overall results are also

comparable with EMCDDA seizure data, with mephedrone (20%), methylene (7%) and MDPV (9%) some of the most commonly seized synthetic cathinones. Furthermore, the United Kingdom had one of the most synthetic cathinone seizures in Europe, implying a high consumption of these NPS [20].

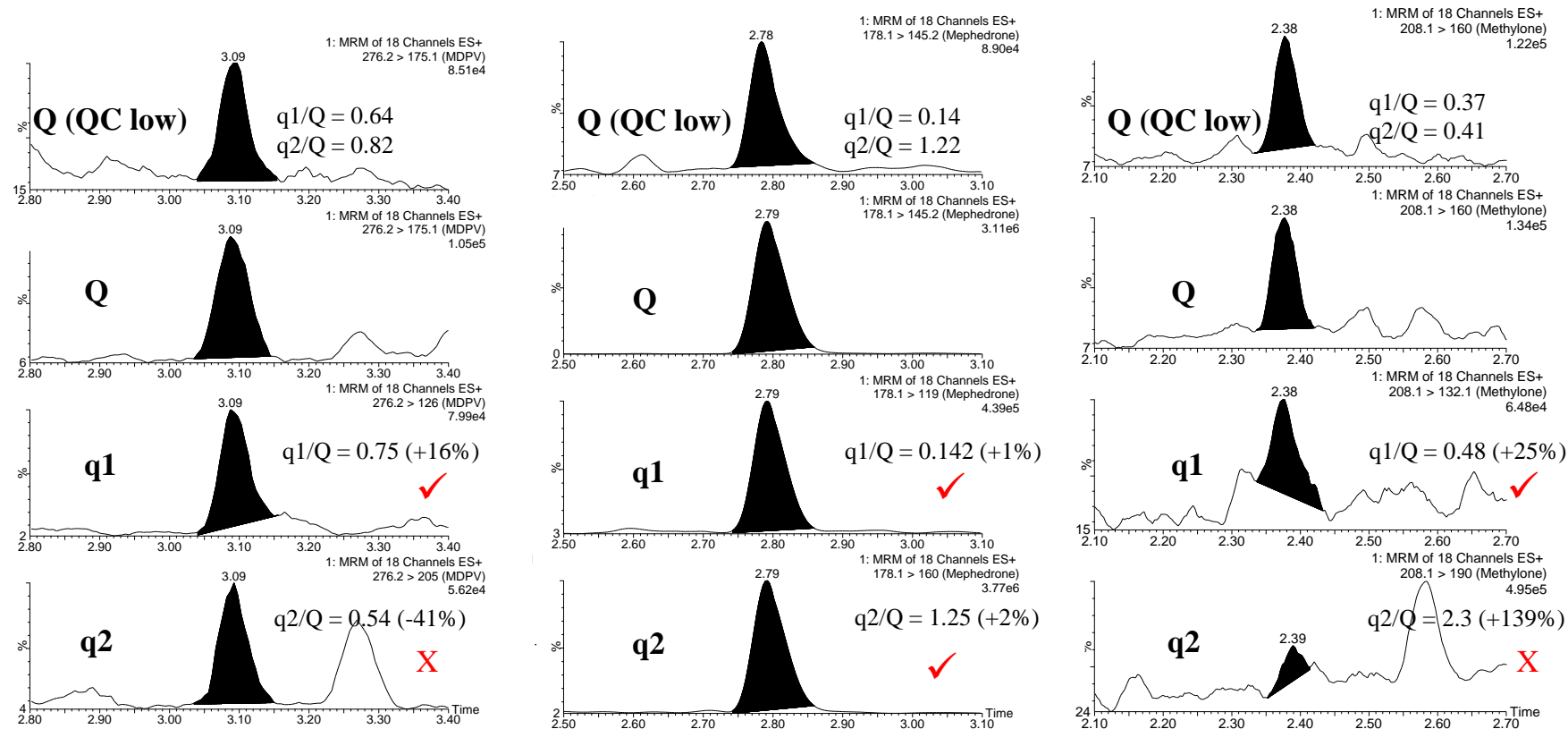


Figure 4: Positive findings (from left to right) of MDPV, mephedrone and methylone in IWW of Milan, South West UK and Copenhagen, respectively. The uppermost transition for each compound is the confirmation (Q) transition in the QC (low, 5 ng L⁻¹), while the lower three are all from the sample. All include three transitions, and the q1/Q ratio, together with the deviation from the QC.

Figure 4 shows the confirmation in three IWW samples of MDPV, mephedrone and methylone. All show the compliance of all the necessary criteria: retention time compatibility with the standard (in the form of the QC in the figure), at least two transitions, and ion ratio within 30% for at least one confirmation transition. It must be noted that the q2/Q ratio for both MDPV and methylone are outside the accepted $\pm 30\%$ ion ratio window, but as the q1/Q ratio for both are within the window, both can be confirmed.

3.6.1 Occurrence of false positives

With quantification methods, it is easy to distinguish positive findings from negative, using strict criteria such as ion ratio ($\pm 30\%$), retention time (± 0.1 min) and at least one additional, confirmatory transition. However, there are some scenarios where the rulings are not so obvious.

For some compounds, only one, the most sensitive, transition may be observed, but may not be confirmed with a subsequent transition, suggesting that the compound could be present, although only at very low concentrations. If criteria are not strictly adhered to, this could lead to false positive findings. This was the case for MDPV in a sample from Zurich (Figure 5). Within this sample, the Q transition was found at the correct retention time, but the two q transitions were between 0.09-0.2 minutes away. This suggests that two different compounds were present in the samples showing the above transitions. Even though the shift in retention time itself is enough to reject the detection of this compound, the ion ratios were still checked. The q1 transition was found within the q/Q ratio threshold (30%). Although the q2 ratio was more than 100% lower than expected, if the sample was not well checked, it could have been erroneously assigned as a false positive on the basis of one q/Q ratio accomplishment.

Figure 5 also shows the curious case of butylone, which seemed to have been detected in an IWW sample from Bristol. As this figure shows, all three transitions can clearly be seen, albeit at a retention time 0.17 minutes lower than in the quality control. Although one q/Q ratio is within the $\pm 30\%$ threshold, the fact that the retention time is greater than 0.10 minutes from the standard is enough evidence to say that this is not butylone. However, with all the same transitions present, there is a possibility that this compound could be related to butylone. Due to the retention time being less than the reference standard, this compound could be a metabolite, which, due to its more polar characteristics would lead it to elute earlier. Alternatively, it could be a related cathinone-derivative, for example ethylone, which has the same fragmentation, and has been found in previous European samples [7]. However, more research is needed to confirm its identity.

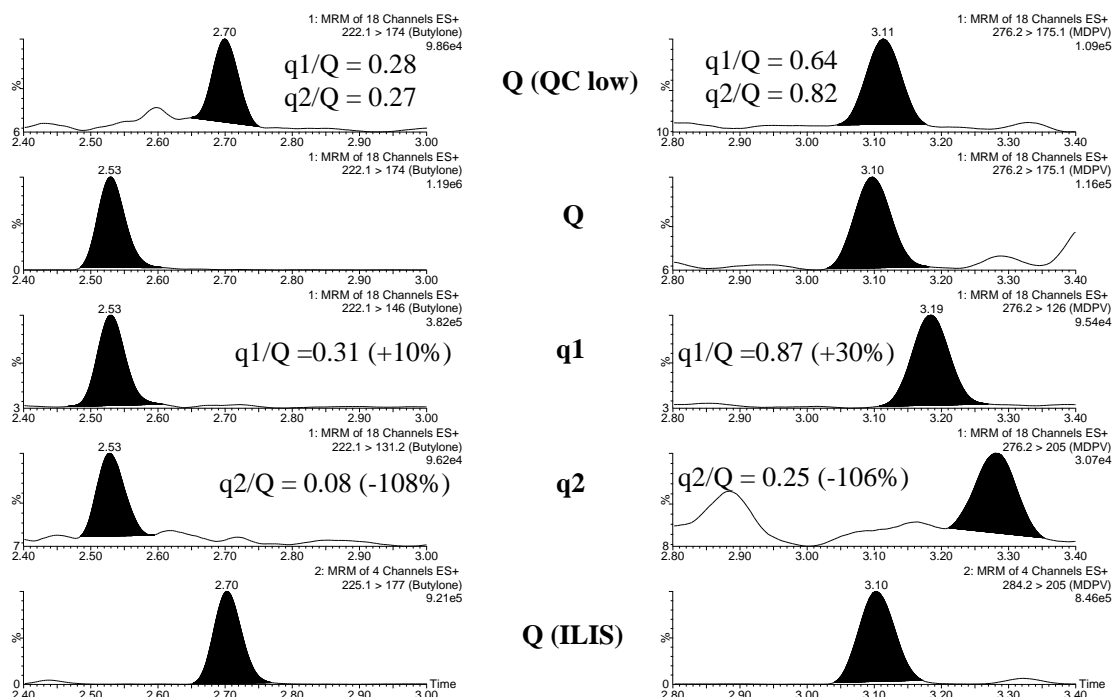


Figure 5: The false positive detections of butylone in Bath (LEFT) and MDPV in Zurich (RIGHT). The uppermost transition is the quantification (Q) transition in the QC (low, 5 ng/L), while the lowermost is the Q of the ILIS (in sample). All transitions are shown, together with the q/Q ratios.

Conclusions

A very sensitive analytical method based on UHPLC-MS/MS has been developed for the determination of ten NPS of the synthetic cathinone and phenethylamine classes in influent wastewater. The overall analytical procedure, based on an off-line SPE step using Oasis MCX cartridges prior to the determination by UHPLC-MS/MS using a triple quadrupole analyzer, has been fully validated, obtaining satisfactory accuracy and precision. Extra emphasis was made on the SPE cartridge, clean-up step and matrix effects. An additional study was also made into the stability of these compounds in five matrices at -20 °C, 4 °C and 20 °C, with advice given on the optimal means of storage.

The method was tested on samples from throughout Europe, with mephedrone, methylone and MDPV able to be identified in various samples. These data support the applicability of the method to influent wastewaters, despite the very low concentrations observed, and also support the previous information available on NPS use in Europe, where these compounds are in agreement with prevalence and seizure data.

In the future, with information on flow data for the IWW plants and metabolic data to correct for excretion, this method could be used to obtain accurate analytical concentrations which subsequently could allow the estimation of NPS consumption within a population.

Contributions

RB, LB, JVS and FH planned and designed the study. LB, JBL, SC, EC, AC, EGL, BKH, JK, AKM, AvN, CO, BGP, PR, NIR, YR, KVT, EZ and PdV organised the collection of the wastewater samples. RB and LB developed the method and RB performed the validation. RB analysed the samples and interpreted the results with contribution from all co-authors. RB, with significant contributions from LB, JVS and FH drafted the manuscript, which was critically revised by all co-authors. All authors are aware of the content, and accept responsibility, for the manuscript.

Acknowledgements

The authors would like to thank all the personnel of the wastewater treatment plants for their collaboration in providing the samples: Sociedad de Fomento Agrícola Castellonense (FACSA, Castellon, Spain), VEAS (Oslo, Norway), Aquiris (Brussels, Belgium), Hoogheemraadschap de Stichtse Rijnlanden (Utrecht, the Netherlands), Wessex Water (South West UK), Metropolitana Milanese (Milan, Italy), Werdhölzli (Zurich, Switzerland) and BIOFOS (Copenhagen, Denmark).

RB, JBL, EC, AC, JK, AKM, PR and NIR acknowledge the European Union for their Early Stage Researcher (ESR) contracts and EGL for her Experienced Researcher (ER) contract as part of the EU-International Training Network SEWPROF (Marie Curie- PEOPLE Grant #317205).

LB acknowledges NPS-Euronet (HOME/2014/JDRUG/AG/DRUG/7086), co-funded by the European Union, for his post-doctoral fellowship. This publication reflects the views only of the authors, and the European Commission cannot be held responsible for any use which may be made of the information contained therein. AvN acknowledges the support of the Flanders Foundation for Research (FWO). EGL is also thankful for the financial support from Generalitat Valenciana (APOSTD/2015, Programa VALi+d).

The financial support of Generalitat Valenciana (Prometeo II 2014/023) and of the Spanish Ministry of Economy and Competitiveness (Project ref CTQ2015-65603) is also acknowledged by the authors of University Jaume I.

References

- [1] UNODC, Global Synthetic Drugs Assessment: Amphetamine-type stimulants and new psychoactive substances, 2014.
- [2] EMCDDA, European Drug Report: Trends and Developments, 2015.
- [3] M. Reid, K. Thomas, New Psychoactive Substances: analysis and site-specific testing, in: S. Castiglioni (Ed.), *Assess. Illicit Drugs Wastewater Adv. Wastewater-Based Drug Epidemiol.*, 2016: pp. 57–65.
- [4] K. V Thomas, L. Bijlsma, S. Castiglioni, A. Covaci, E. Emke, R. Grabic, et al., Comparing illicit drug use in 19 European cities through sewage analysis, *Sci. Total Environ.* 432 (2012) 432–439. doi:10.1016/j.scitotenv.2012.06.069.
- [5] C. Ort, A.L.N. van Nuijs, J.-D. Berset, L. Bijlsma, S. Castiglioni, A. Covaci, et al., Spatial differences and temporal changes in illicit drug use in Europe quantified by wastewater analysis, *Addiction.* 109 (2014) 1338–1352. doi:10.1111/add.12570.
- [6] C. Chen, C. Kostakis, R.J. Irvine, J.M. White, Increases in use of novel synthetic stimulant are not directly linked to decreased use of 3,4-methylenedioxy-N-methylamphetamine (MDMA), *Forensic Sci. Int.* 231 (2013) 278–283. doi:10.1016/j.forsciint.2013.06.007.
- [7] J. Kinyua, A. Covaci, W. Maho, A.-K. McCall, H. Neels, A.L.N. van Nuijs, Sewage-based epidemiology in monitoring the use of new psychoactive substances: Validation and application of an analytical method using LC-MS/MS, *Drug Test. Anal.* 7 (2015) 812–818. doi:10.1002/dta.1777.
- [8] I. Senta, I. Krizman, M. Ahel, S. Terzic, Multiresidual analysis of emerging amphetamine-like psychoactive substances in wastewater and river water, *J. Chromatogr. A.* 1425 (2015) 204–212. doi:10.1016/j.chroma.2015.11.043.
- [9] V.L. Borova, P. Gago-Ferrero, C. Pistos, N.S. Thomaidis, Multi-residue determination of 10 selected new psychoactive substances in wastewater samples by liquid chromatography–tandem mass spectrometry, *Talanta.* 144 (2015) 592–603. doi:10.1016/j.talanta.2015.06.080.
- [10] A. Kankaanpää, K. Ariniemi, M. Heinonen, K. Kuoppasalmi, T. Gunnar, Use of illicit stimulant drugs in Finland: A wastewater study in ten major cities, *Sci. Total Environ.* 487 (2014) 696–702. doi:10.1016/j.scitotenv.2013.11.095.
- [11] A.L.N. van Nuijs, A. Gheorghe, P.G. Jorens, K. Maudens, H. Neels, A. Covaci, Optimization, validation, and the application of liquid chromatography-tandem mass spectrometry for the analysis of new drugs of abuse in wastewater., *Drug Test. Anal.* (2013) 861–867. doi:10.1002/dta.1460.
- [12] M.J. Reid, L. Derry, K. V. Thomas, Analysis of new classes of recreational drugs in sewage: Synthetic

- cannabinoids and amphetamine-like substances, *Drug Test. Anal.* 6 (2014) 72–79.
doi:10.1002/dta.1461.
- [13] E.M. Mwenesongole, L. Gautam, S.W. Hall, J.W. Waterhouse, M.D. Cole, Simultaneous detection of controlled substances in waste water, *Anal. Methods.* 5 (2013) 3248–3254. doi:10.1039/c3ay40655e.
- [14] J.A. Baz-Lomba, M.J. Reid, K. V. Thomas, Target and suspect screening of psychoactive substances in sewage-based samples by UHPLC-QTOF, *Anal. Chim. Acta.* 914 (2016) 81–90.
doi:10.1016/j.aca.2016.01.056.
- [15] E. Castrignanò, A. Lubben, B. Kasprzyk-Hordern, Enantiomeric profiling of chiral drug biomarkers in wastewater with the usage of chiral liquid chromatography coupled with tandem mass spectrometry, *J. Chromatogr. A.* 1438 (2016) 84–99. doi:10.1016/j.chroma.2016.02.015.
- [16] Ó.J. Pozo, J. V. Sancho, M. Ibáñez, F. Hernández, W.M. a. Niessen, Confirmation of organic micropollutants detected in environmental samples by liquid chromatography tandem mass spectrometry: Achievements and pitfalls, *TrAC Trends Anal. Chem.* 25 (2006) 1030–1042.
doi:10.1016/j.trac.2006.06.012.
- [17] S. Elliott, J. Evans, A 3-year review of new psychoactive substances in casework, *Forensic Sci. Int.* 243 (2014) 55–60. doi:10.1016/j.forsciint.2014.04.017.
- [18] V. Uralets, S. Rana, S. Morgan, W. Ross, Testing for designer stimulants: Metabolic profiles of 16 synthetic cathinones excreted free in human urine, *J. Anal. Toxicol.* 38 (2014) 233–241.
doi:10.1093/jat/bku021.
- [19] M. Ibáñez, J. V. Sancho, L. Bijlsma, A.L.N. van Nuijs, A. Covaci, F. Hernández, Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances, *TrAC Trends Anal. Chem.* 57 (2014) 107–117.
doi:10.1016/j.trac.2014.02.009.
- [20] EMCDDA, New psychoactive substances in Europe, *EU Early Warn. Syst.* (2015) 12.
doi:10.2810/372415.
- [21] SANCO, Guidance document on analytical quality control and validation procedures for pesticides residues in food and feed, *SANCO /12571/2013.* (2013).
http://ec.europa.eu/food/plant/pesticides/guidance_documents/docs/qualcontrol_en.pdf.
- [22] L. Bijlsma, J. V Sancho, E. Pitarch, M. Ibanez, F. Hernandez, Simultaneous ultra-high-pressure liquid chromatography-tandem mass spectrometry determination of amphetamine and amphetamine-like stimulants, cocaine and its metabolites, and a cannabis metabolite in surface water and urban wastewater, *J. Chromatogr. a.* 1216 (2009) 3078–3089. doi:10.1016/j.chroma.2009.01.067.

- [23] H.T. Kamata, N. Shima, K. Zaitso, T. Kamata, A. Miki, M. Nishikawa, et al., Metabolism of the recently encountered designer drug, methylone, in humans and rats, *Xenobiotica*. 36 (2006) 709–723. doi:10.1080/00498250600780191.
- [24] M.R. Meyer, P. Du, F. Schuster, H.H. Maurer, Studies on the metabolism of the α -pyrrolidinophenone designer drug methylenedioxy-pyrovalerone (MDPV) in rat and human urine and human liver microsomes using GC-MS and LC-high-resolution MS and its detectability in urine by GC-MS, *J. Mass Spectrom.* 45 (2010) 1426–1442. doi:10.1002/jms.1859.
- [25] M. Mardal, M.R. Meyer, Studies on the microbial biotransformation of the novel psychoactive substance methylenedioxy-pyrovalerone (MDPV) in wastewater by means of liquid chromatography-high resolution mass spectrometry/mass spectrometry, *Sci. Total Environ.* 493 (2014) 588–595. doi:10.1016/j.scitotenv.2014.06.016.
- [26] M. Ibáñez, Ó.J. Pozo, J. V. Sancho, T. Orengo, G. Haro, F. Hernández, Analytical strategy to investigate 3,4-methylenedioxy-pyrovalerone (MDPV) metabolites in consumers' urine by high-resolution mass spectrometry, *Anal. Bioanal. Chem.* 408 (2016) 151–164. doi:10.1007/s00216-015-9088-1.
- [27] A.T. Caspar, A.G. Helfer, J.A. Michely, V. Auwarter, S.D. Brandt, M.R. Meyer, et al., Studies on the metabolism and toxicological detection of the new psychoactive designer drug 2-(4-iodo-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine (25I-NBOMe) in human and rat urine using GC-MS, LC-MS, and LC-HR-MS/MS, *Anal Bioanal Chem.* (2015) 6697–6719. doi:10.1007/s00216-015-8828-6.
- [28] F.Y. Lai, C. Erratico, J. Kinyua, J.F. Mueller, A. Covaci, A.L.N. van Nuijs, Liquid chromatography-quadrupole time-of-flight mass spectrometry for screening in vitro drug metabolites in humans: investigation on seven phenethylamine-based designer drugs, *J. Pharm. Biomed. Anal.* 114 (2015) 355–375. doi:10.1016/j.jpba.2015.06.016.
- [29] J. Kinyua, N. Negreira, M. Ibáñez, L. Bijlsma, F. Hernández, A. Covaci, et al., A data-independent acquisition workflow for qualitative screening of new psychoactive substances in biological samples, *Anal. Bioanal. Chem.* 407 (2015) 8773–8785. doi:10.1007/s00216-015-9036-0.
- [30] J.L. Poklis, C.R. Nanco, M.M. Troendle, C.E. Wolf, A. Poklis, Determination of 4-bromo-2,5-dimethoxy-N-[(2-methoxyphenyl)methyl]-benzeneethanamine (25B-NBOMe) in serum and urine by high performance liquid chromatography with tandem mass spectrometry in a case of severe intoxication, *Drug Test. Anal.* 6 (2014) 764–769. doi:10.1002/dta.1522.
- [31] M.H. Tang, C.K. Ching, M.S. Tsui, F.K. Chu, T.W. Mak, Two cases of severe intoxication associated

- with analytically confirmed use of the novel psychoactive substances 25B-NBOMe and 25C-NBOMe, *Clin Toxicol.* 52 (2014) 561–565. doi:10.3109/15563650.2014.909932.
- [32] D.R. Baker, B. Kasprzyk-Hordern, Critical evaluation of methodology commonly used in sample collection, storage and preparation for the analysis of pharmaceuticals and illicit drugs in surface water and wastewater by solid phase extraction and liquid chromatography–mass spectrometry, *J. Chromatogr. A.* 1218 (2011) 8036–8059. doi:10.1016/j.chroma.2011.09.012.
- [33] L. Bijlsma, E. Beltrán, C. Boix, J. V Sancho, F. Hernández, Improvements in analytical methodology for the determination of frequently consumed illicit drugs in urban wastewater, *Anal. Bioanal. Chem.* 406 (2014) 4261–4272. doi:10.1007/s00216-014-7818-4.
- [34] R. Bade, N.I. Rousis, L. Bijlsma, E. Gracia-Lor, S. Castiglioni, J. V. Sancho, et al., Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS, *Anal. Bioanal. Chem.* 407 (2015) 8979–8988. doi:10.1007/s00216-015-9063-x.
- [35] B.K. Matuszewski, M.L. Constanzer, C.M. Chavez-Eng, Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS, *Anal. Chem.* 75 (2003) 3019–3030. doi:10.1021/ac020361s.
- [36] A.-K. McCall, R. Bade, J. Kinyua, F.Y. Lai, P.K. Thai, A. Covaci, et al., Critical review on the stability of illicit drugs in sewers and wastewater samples, *Water Res.* 88 (2016) 933–947. doi:10.1016/j.watres.2015.10.040.

Supporting Information**Table S1:** *Filtration recovery for all compounds using vacuum filter (%)*

Compound	Natural pH	pH = 2
N-ethylcathinone	98	106
MDPV	55	71
Mephedrone	97	102
Methylone	88	104
Butylone	84	104
Methedrone	90	102
Naphyrone	1	4
25-B-NBOMe	1	1
25-I-NBOMe	0	1
25-C-NBOMe	1	1

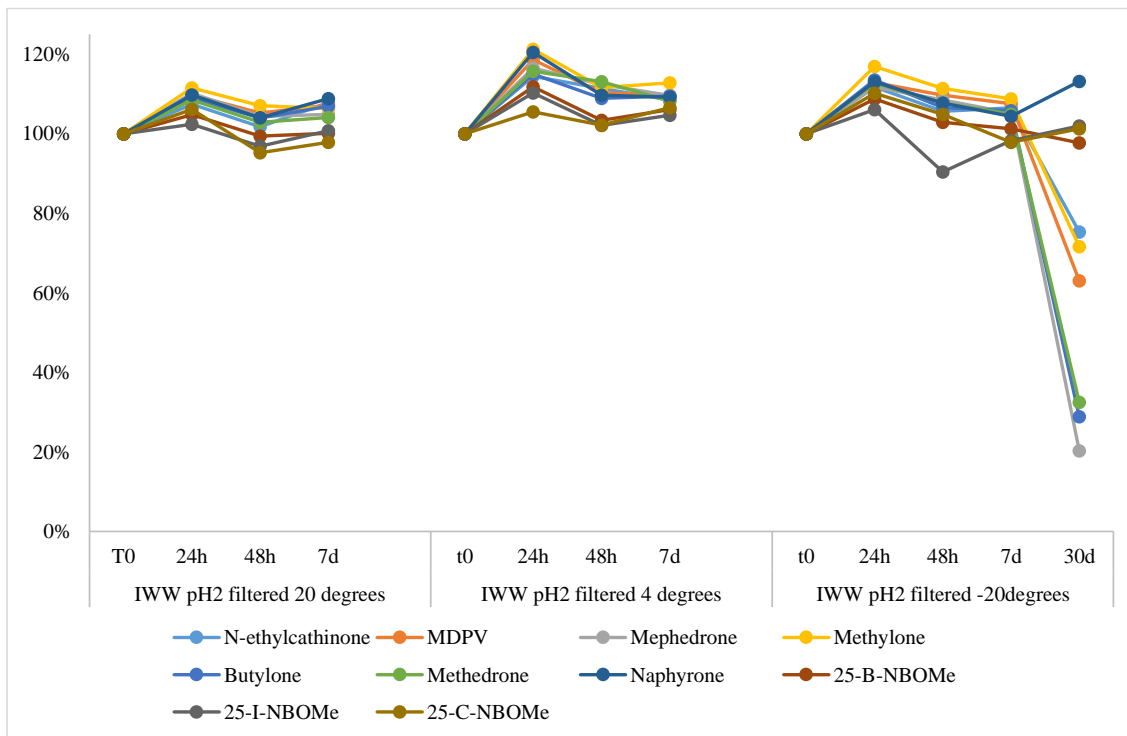


Figure S1: Stability of all compounds in filtered IWW (pH2) at 20°C (LEFT), 4°C (MIDDLE) and -20°C (RIGHT)

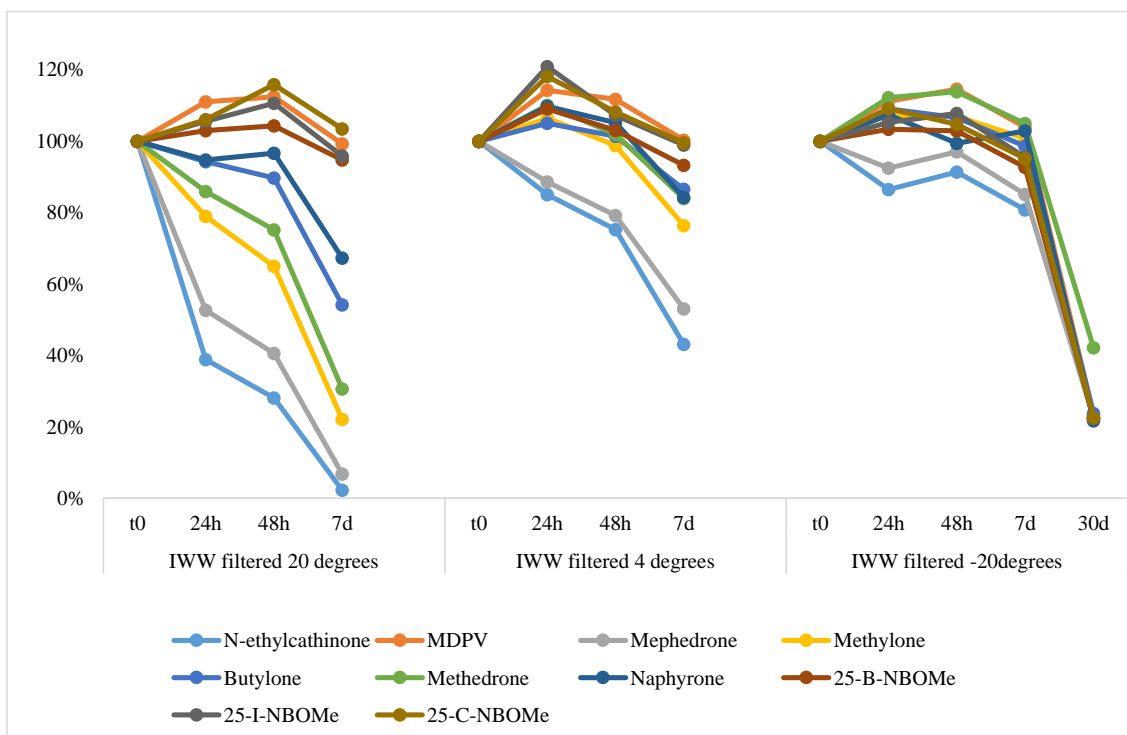


Figure S2: Stability of all compounds in filtered IWW (natural pH) at 20°C (LEFT), 4°C (MIDDLE) and -20°C (RIGHT)

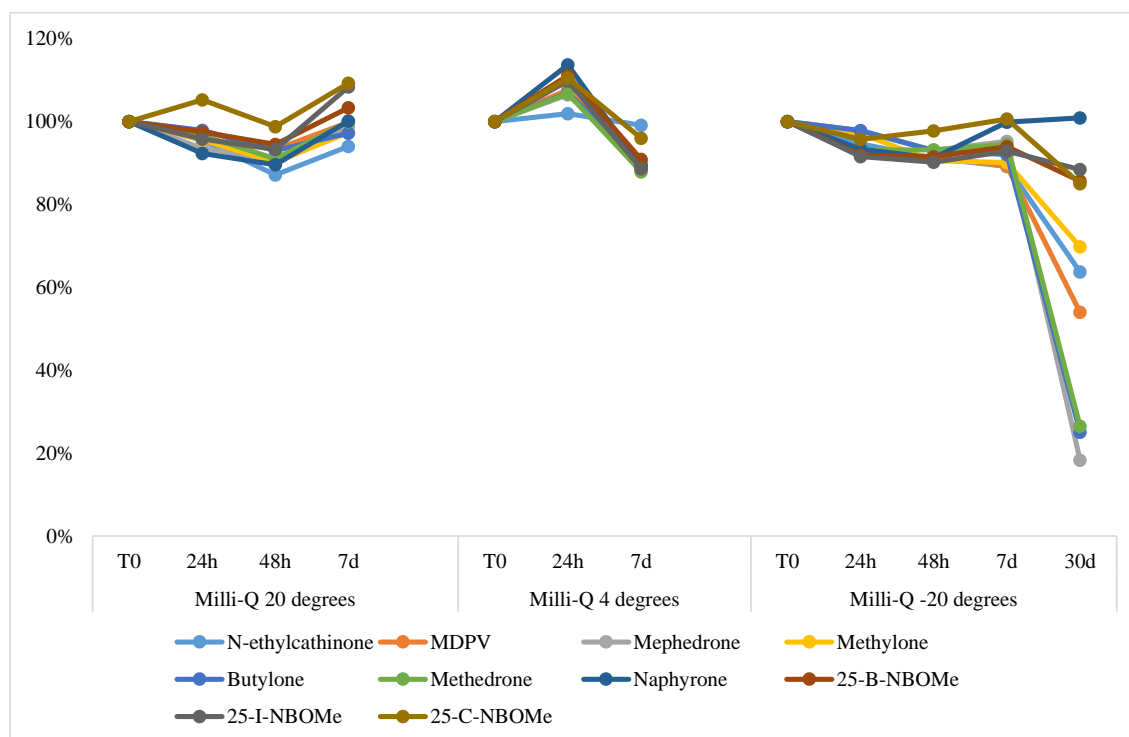


Figure S3: Stability of all compounds in Milli-Q water at 20°C (LEFT), 4°C (MIDDLE) and -20°C (RIGHT)

4.3 Discussion of Results

Within the realm of WBE, illicit drugs (especially cocaine, methamphetamine, amphetamine, MDMA, and cannabis) have been the focus. The success of this approach has led researchers to explore further biomarkers in fields other than illicit drugs, including F₂-isoprostanes, alcohol, caffeine, nicotine, phosphodiesterase type V inhibitors and pesticides. In the past five years, there has been a rapid growth in NPS, as 'legal' alternatives to the conventional illicit drugs outlined above. With little toxicological knowledge of these compounds, governments and regulators have been quick to illegalise these compounds. However, slight modifications can be made to get around the legislation, leading to more NPS hitting the market. In spite of the legislation, and the new NPS continuously reaching the market, some NPS have proven more popular than others (even after being deemed illegal), such as mephedrone (in Europe) and MDPV and methylene (mainly in the United States).

Several research groups have developed WBE methods for various NPS (Borova et al., 2015; Kinyua et al., 2015a; Senta et al., 2015), while our group has previously analysed legal highs – in the form of herbal blends, pills, powders, drinkable solutions and truffles – to identify the NPS inside (Ibáñez et al., 2014). This information, combined with that from recent toxicological studies (Caspar et al., 2015; Kinyua et al., 2015b; Mardal and Meyer, 2014), led to the ten NPS being chosen for this study: N-ethylcathinone, butylone, methylene, naphyrone, MDPV, mephedrone, methedrone, 25-B-NBOMe, 25-C-NBOMe and 25-I-NBOMe.

Only parent compounds were selected for this study. This is based on the previous metabolism studies made on these and related NPS in human and rat urine (Caspar et al., 2015; Ibáñez et al., 2016; Kamata et al., 2006; Mardal and Meyer, 2014; Meyer et al., 2010; Uralets et al., 2014), which showed that in spite of the number of metabolites identified, the parent compound could be detected, thereby giving credence to their inclusion as a biomarker for drug consumption. Furthermore, it has been noted that the metabolic patterns of selected phenethylamine-based designer drugs have indicated rather slow metabolism rates, with parent drugs being the key biomarkers of the drug consumption (Lai et al., 2015; Senta et al., 2015), somewhat expected due to their structural similarity to amphetamine and MDMA, for both of which the parent drug is used as the biomarker for drug consumption.

Matters are further complicated with information relating to routes of administration, dosage levels, habits of consumption and individual metabolism lacking. This could have a big impact on the most suitable biomarker, as the excretion rate can vary according to the route of administration and the

frequency of use of a substance (Castiglioni et al., 2016, 2013). Although back-calculation to estimate consumption is not (yet) the goal of identifying NPS in wastewater, knowledge of the most highly excreted metabolites could allow more of these NPS to be found, and can be used for future WBE methods.

With the conventional WBE drugs, excretion rates of the biomarkers used range from 0.6% (THC-COOH for cannabis) to 65 % (MDMA) (Burgard et al., 2014; Castiglioni and Gracia-Lor, 2016). The fact that these compounds are still utilised as the most suitable biomarkers despite the small excretion rates in some cases, combined with the findings of the aforementioned metabolism studies, leads the parent drugs to being a plausible biomarker of consumption.

When developing this method, the first parameters to be optimised were those related to the mass spectrometry measurement. As with other methods incorporating these basic compounds, only positive ionisation mode was tested, with the $[M+H]^+$ used as the precursor ion for all compounds for both quantification and confirmation transitions. Initially, all analytes were injected as a mix, with optimisation of the cone voltage made concurrently, with the voltage being increased from 10V-50V at 10V steps (example of methedrone in **Figure 4.4**). If the cone voltage deemed to have the greatest sensitivity was at 50V, higher cone voltages were applied (as in the cases of methylone and naphyrone), to see whether greater sensitivity could be achieved. Once the optimal cone voltage had been found, the collision energy was optimised for each compound individually, using the optimised cone voltage and the collision energy being increased from 10eV-50eV at 10eV increments. All transitions (i.e. precursor ion > product ion) were noted, along with their optimal collision energy and areas (to determine the sensitivity).

In general, the most sensitive transition was used for quantification. The only exception was methylone where it corresponded to the non-specific loss of water, thus the 178.1>145.2 transition was used for quantification instead. For confirmation, the next most sensitive transition was selected, with the loss of water again overlooked due to its low specificity. However, for mephedrone, a loss of water (178.1>160.0) was used for confirmation due to the low sensitivity of other transitions.

The initial UHPLC optimisation was based on a previous method of our group (Bijlsma et al., 2014). A mobile phase of pure milli-Q water (H₂O) and methanol (MeOH) was first used, together with a mixed solution of the analytes (100ppb). However, due to the sensitivity of the instrument, the column was overloaded at such a high concentration, so subsequent injections were made at 10ppb and 1ppb. Furthermore, without any additives, no clear chromatographic peaks could be observed. Formic acid

(FA) and ammonium formate (NH_4Ac) at differing concentrations were then added in an attempt to improve the chromatographic peaks (**Figure 4.4**). As seen, there is a massive improvement as soon as FA is added to the aqueous phase.

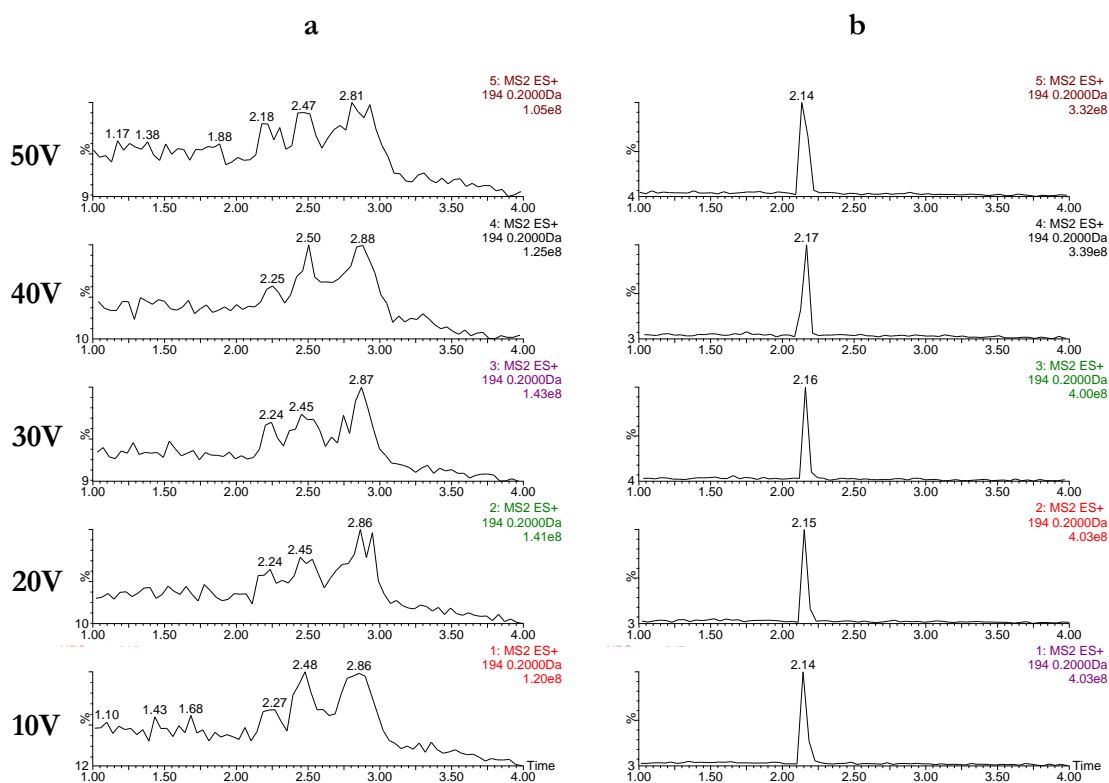


Figure 4.4: Cone voltage and UHPLC Optimisation of Methedrone: *a* = 10ppb methedrone, mobile phase of H_2O and MeOH only; *b* = 10ppb methedrone, mobile phase of H_2O + 0.01%FA and MeOH.

Cone voltage is shown on the left.

When tweaking the mobile phase additives, it was found that the compounds most susceptible to peak broadening were those eluting later: naphyrone and the NBOMe compounds. While the addition of FA and NH_4Ac to the aqueous mobile phase did improve the peak shapes of all compounds, the addition of FA to the organic mobile phase improved the sensitivity of all compounds, with up to a two-fold increase (**Figure 4.5**). Thus, a mobile phase of water (with 0.01% formic acid and 5mM ammonium acetate) and methanol (with 0.01% formic acid) was selected.

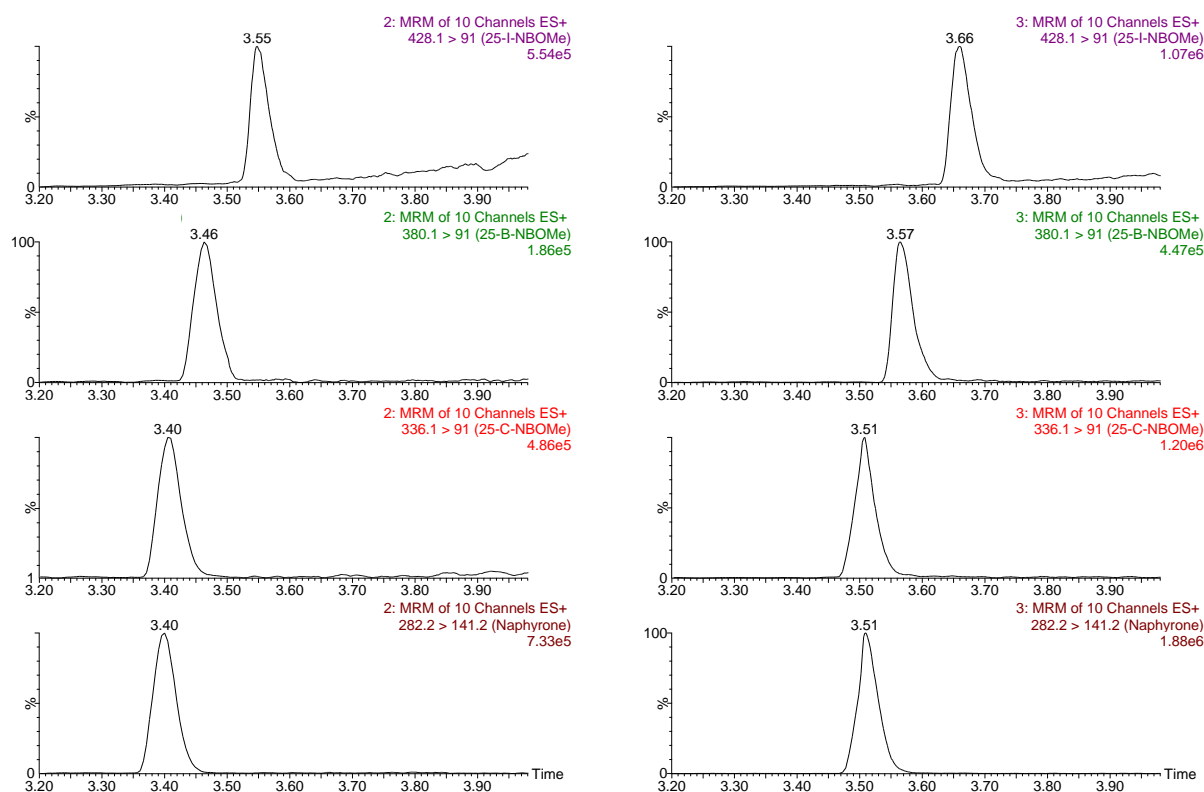


Figure 4.5: UHPLC Optimisation: LEFT: H₂O + 0.01% FA + 5mMol NH₄Ac; MeOH, RIGHT: H₂O + 0.01% FA + 5mMol NH₄Ac; MeOH + 0.01% FA
Compounds, from top to bottom: 25-I-NBOMe, 25-B-NBOMe, 25-C-NBOMe, naphyrone

Two SPE cartridges were tested for both sample treatment purposes: Oasis HLB and Oasis MCX. Oasis HLB is a hydrophilic-lipophilic-balanced, water wettable, reversed phase material, commonly used as a generic cartridge as it is ideal for acidic, basic and neutral analytes. Oasis MCX is a mixed mode cation-exchange sorbent, which is selective for bases. Both consist of a copolymer synthesised from a polar divinylbenzene and polar N-vinylpyrrolidone. HLB is only synthesised from these two monomers and neutral. MCX has extra sulfonic acid functional groups (1.01mmol/g) implemented at a fixed location and contain permanently charged groups at any relevant pH (Bauerlein et al., 2012). All of the compounds in this method were basic phenethylamine- and cathinone-derivatives, so the MCX was expected to be superior, or at least being more specific. In order to compare the two cartridges, recovery experiments were made, whereby the samples were spiked before and after extraction. **Figure 4.6** shows the comparison between the two procedures at 10ppb in IWW, with the MCX giving recoveries between 80-100% for all compounds, while, using HLB, both butylone and methedrone had recoveries less than 80%. This led to MCX being chosen as the preferred sample treatment method, and further optimisation of the elution and drying was made to improve the

recoveries. However, as shown in Figure 2 of **Scientific Article 7**, the recoveries for these alternative steps led to recoveries greater than 120%.

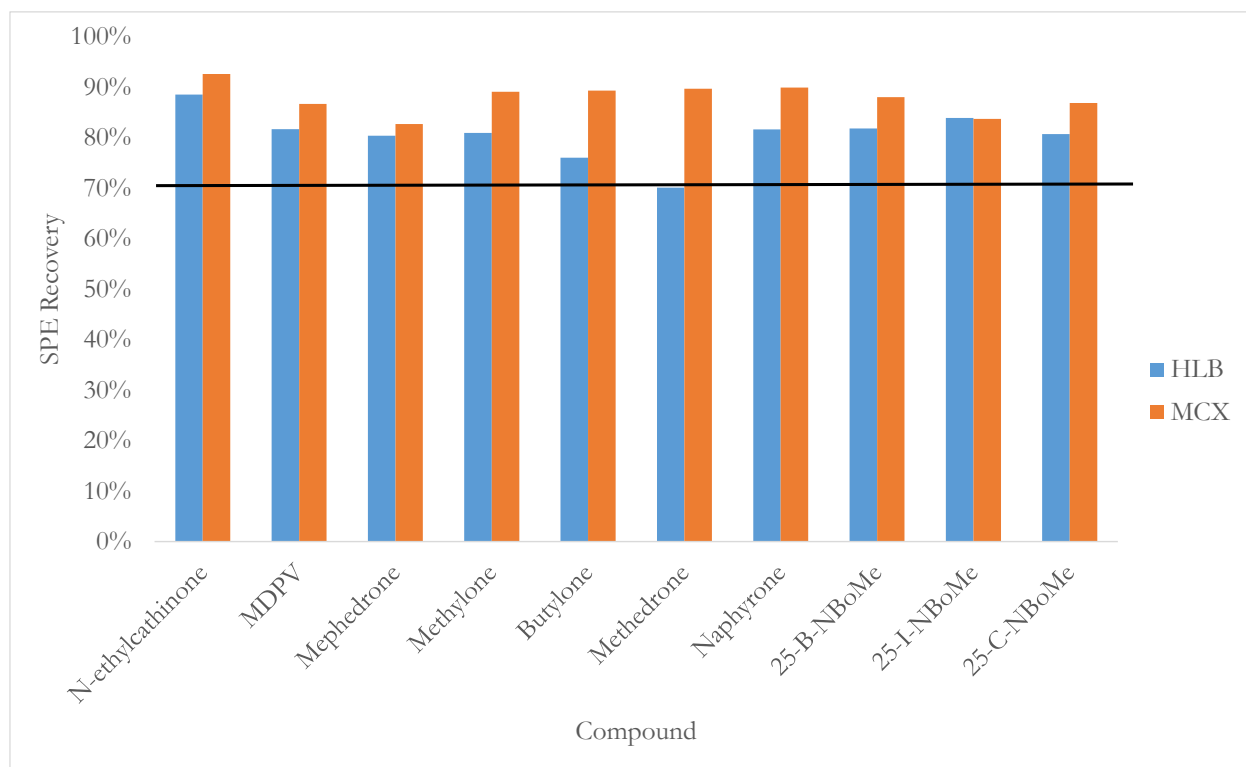


Figure 4.6: Comparison of SPE recoveries of Oasis HLB and Oasis MCX

Matrix effects, extraction efficiency (absolute recovery) and process efficiency (relative recovery) were also tested for the MCX cartridges, based on the work of Matuszewski et al. (Matuszewski et al., 2003), where three different sets of samples were prepared. Set 1 consisted of a mixed standard solution in solvent, including ILIS. Set 2 consisted of IWW samples, which had previously been extracted, and the SPE eluates spiked with the mixed standard and ILIS solutions. Finally, in set 3, IWW samples were first spiked at the same concentrations as the previous sets and then extracted. All sets were performed in triplicate and averaged to perform calculations. Extraction efficiency was calculated by comparing the peak areas of set 2 and set 3; matrix effects by comparing set 2 and set 1, and total process efficiency by comparing set 3 and set 1, as in the calculations below, with **Table 4.1** showing the results of these experiments:

$$\text{Extraction Efficiency / Absolute Recovery (\%)} = \frac{\text{average peak area (Set 2)}}{\text{average peak area (Set 3)}} \times 100$$

Corrected Extraction Efficiency/ Absolute Recovery (%) =

$$\frac{\text{average peak area (Set 2, ILIS)} / \text{average peak area (Set 3, ILIS)}}{\text{average peak area (Set 2)} / \text{average peak area (Set 3)}} \times 100$$

Matrix effects (%) = $\frac{\text{average peak area (Set 1)}}{\text{average peak area (Set 2)}} \times 100$; Matrix effects (%) – 1 = matrix enhancement (i.e.

all positive figures denote matrix enhancement, all negative figures matrix suppression)

Corrected Matrix effects (%) = $\frac{\text{average peak area (Set 1, ILIS)} / \text{average peak area (Set 2, ILIS)}}{\text{average peak area (Set 1)} / \text{average peak area (Set 2)}} \times 100$

Process Efficiency/Relative Recovery (%): = $\frac{\text{average peak area (Set 3)}}{\text{average peak area (Set 1)}} \times 100$

Corrected Process Efficiency/Relative Recovery (%) =

$$\frac{\text{average peak area (Set 3, ILIS)} / \text{average peak area (Set 1, ILIS)}}{\text{average peak area (Set 3)} / \text{average peak area (Set 1)}} \times 100$$

Table 4.1: Absolute recovery, matrix effects and relative recovery (%) for all compounds after SPE with MCX cartridges

Compound	Absolute recovery	Absolute recovery (corrected ^b)	Matrix effects ^a	Matrix Effects (corrected ^b)	Relative recovery	Relative recovery (corrected ^b)
N-ethylcathinone	83	102	73	3	23	101
MDPV	91	102	52	-7	44	91
Mephedrone	85	103	74	-9	22	88
Methylone	84	104	70	-6	25	91
Butylone	89	99	71	-6	26	95
Methedrone	87	97	69	-10	27	94
Naphyrone	87	99	38	-10	54	91
25-B-NBoMe	89	96	40	-4	53	100
25-I-NBoMe	86	97	34	-6	57	96
25-C-NBoMe	90	103	50	-7	45	90

^a: for matrix effects, a negative value (-) denotes matrix suppression, while positive values indicate matrix enhancement.

^b: corrected with ILIS

The use of ILIS was imperative, as they allowed the correction of both matrix effects and losses associated with the sample preparation. In this respect, a deuterated analogue of each individual analyte would be an ideal internal standard, as it should show an identical behaviour to the sample analyte in sample pretreatment, chromatography and ionisation. All but two analytes had a deuterated analyte as an internal standard, but alternative ILIS were used for N-ethylcathinone and methedrone (methylone-d₃ and butylone-d₃, respectively) owing to their ability to correct for matrix effects and their similar retention times.

A direct injection method was also attempted, as a comparison to the applied SPE method. One clear advantage of direct injection is that SPE is not needed, thereby cutting sample pretreatment times considerably, and avoiding potential losses or contaminations that may occur during this step. However, without any preconcentration, there is a chance that the levels able to be reached by direct injection are not as low as those that incorporate SPE. Furthermore, there is also a high chance of matrix interferences, leading to signal suppression or enhancement, which can make validation and quantification difficult. Normally, direct injection in this field involves large volume injection (i.e. from 50-900 µL), generally for “cleaner” samples such as drinking water and surface water; however, an effort was made to use an even lower injection volume, of 3 µL, to reduce matrix effects further. It was found that the matrix effects led to signal enhancement for all the compounds (**Table 4.2**), and were influenced by their retention time, with later eluting compounds having greater matrix effects (two- to three-fold higher) than the earlier eluting compounds. This could be as a result of the gradient and the “washing” of the column at the later stages of the chromatographic run. Acceptable matrix effects were observed for the first five eluting compounds (from Methylone to Mephedrone), while they lead up to 2.5-fold signal for the last eluting compound 25-B-NBoMe. More research is needed in this area in order to facilitate direct injection of samples in the future (see Future Perspectives).

Table 4.2: Matrix effects (%) of all compounds with direct injection method

Compound	Retention Time	Matrix Effects ^a
Methylone	2.07	25
N-ethylcathinone	2.14	4
Methedrone	2.31	14
Butylone	2.44	11
Mephedrone	2.53	7
MDPV	2.88	86
25-C-NBoMe	3.50	196
Naphyrone	3.51	212
25-B-NBoMe	3.56	240

^aMatrix effects of 0% indicate no suppression or enhancement, while all positive values indicate enhancement

Within the development of a direct injection method, a study was made into the recoveries of the compounds after filtration, both using several syringe filters (for direct injection) and a vacuum filter normally used in our laboratory prior to both qualitative and quantitative methods of wastewater (Bijlsma et al., 2014; Hernández et al., 2015). Filtration experiments were performed using IWW at both natural pH and acidified to pH2, after centrifugation (**Tables 4.3** and **4.4**). Seven syringe filters were tested (0.2µm Nylon; 25mm (Sarorius, only at natural pH), 0.22 µm PTFE; 13mm (Membrane solutions (Mieron)), 0.45µm Nylon; 17mm (Nalgene), 0.45 µm Nylon; 13mm (Scharlau), 0.2 µm Polypropylene; 13mm (Whatman), 0.2µm PTFE ;13mm (Teknokroma) and 0.45µm PTFE; 13mm (Teknokroma)) and one vacuum filter (0.45µm mixed cellulose ester membrane filters (Whatman, Dassel, Germany). For all experiments, the wastewater sample was spiked at 2.5ppb with mixed standards prior to filtration, while 2ppb (final concentration) of the ILIS was added after, to correct for matrix effects. 1mL of IWW was collected after filtration, in triplicate.

Of all of the filters investigated, only one (0.45µm Nylon, 13mm) did not retain any compounds in the filtration of IWW at natural pH, while at pH2, all compounds passed all syringe filters. At natural pH, the NBOME compounds (especially 25-I-NBOME) were particularly retained on the filters, recording recoveries less than 50% for the majority of the filters (**Table 4.3**)

Table 4.3: Filtration recoveries (%) of all compounds at natural pH

Compound	Membrane 0.45µm (vacuum)	Nylon 0.20 µm (25mm)	PTFE 0.22µm (13mm)	Nylon 0.45µm (17mm)	Nylon 0.45µm (13mm)	Polypropylene 0.2 µm (13mm)	PTFE 0.2µm (13mm)	PTFE 0.45µm (13mm)
N-ethylcathinone	98	96	100	92	119	97	123	114
MDPV	55	91	97	70	121	67	95	100
Mephedrone	97	99	118	106	140	93	106	105
Methylone	88	96	104	101	125	99	103	102
Butylone	84	90	98	86	115	81	102	100
Methedrone	90	82	10%	86	115	90	104	102
Naphyrone	1	67	75	67	106	3	64	83
25-B-NBoMe	1	26	28	39	101	2	33	64
25-I-NBoMe	0	0	9	18	91	0	19	50
25-C-NBoMe	1	39	42	55	110	2	40	69

One interesting aspect of this experiment was the effectiveness of the vacuum filter that is commonplace in qualitative and quantitative studies in our laboratory. It was found that at both natural pH and pH2, four compounds (naphyrone and the three NBoMe compounds) were highly retained (less than 5% remaining), with MDPV also being somewhat retained (29-45%) on the filter. This result suggests that screening methods performed using this filter will always be missing these NPS, which could result in their being erroneously qualified as not detected. It must be noted that because of this result, none of the IWW samples for the SPE method were filtered, solely centrifuged. The filter finally chosen for direct injection was the Nylon 13mm, 0.45µm, which had recoveries of 98-105% for all compounds at pH2 and was the only filter to have respectable recoveries for all compounds at natural pH (91-140%).

Table 4.4: Filtration recovery (%) of all compounds at pH 2

Compound	Membrane 0.45µm (vacuum)	PTFE 0.22µm (13mm)	Nylon 0.45µm (17mm)	Nylon 0.45µm (13mm)	Polypropylene 0.2 µm (13mm)	PTFE 0.2µm (13mm)	PTFE 0.45µm (13mm)
N-ethylcathinone	106	97	97	101	96	91	98
MDPV	71	100	89	101	101	103	103
Mephedrone	102	102	101	105	97	100	99
Methylone	104	103	102	104	101	98	100
Butylone	104	101	98	102	92	96	97
Methedrone	102	100	97	103	97	100	101
Naphyrone	4	85	89	95	92	99	102
25-B-NBoMe	1	74	87	99	81	104	99
25-I-NBoMe	1	53	82	98	64	102	83
25-C-NBoMe	1	79	87	98	87	105	101

Tables 1 and 2 in **Scientific Article 7** show the results of the final validation of the SPE method (n=5), with precision (RSD), recovery, LOD, LOQ and linearity. The limit of quantification objective (LOQ) was taken as the lowest level for which the method was fully validated using spiked samples with satisfactory recovery (between 70-120%) and precision (relative standard deviation (RSD) \leq 20%). Two LOQ validations (1ng/L and 5ng/L) were made to provide a more accurate measure, with MDPV, naphyrone and the NBOME compounds all having LOQs as low as 1ng/L. Recoveries were acceptable for all compounds, ranging between 71-113% (at 5ng/L) and 85-122% (and 1ng/L), and the precision (in terms of RSD) was below 20% for all compounds. Linearity was studied for all compounds. Two calibration curves were made: 0.5-25ng/L (five points) and 1-100ng/L (6 points), the first for those compounds with a lower LOQ (1ng/L). Correlation coefficients were greater than 0.99 for all compounds, except naphyrone and the NBOME compounds, which had values greater than 0.96.

A selection of samples from throughout Europe was investigated for NPS using this method. In total, 17 samples were investigated from Bristol, United Kingdom (1 x weekday, 1 x weekend); Brussels, Belgium (1 x weekday, 1 x weekend); Castellon, Spain (1 x weekend) Oslo, Norway (1 x weekday, 1 x weekend); Copenhagen, Denmark (1 x weekday, 2 x weekend); Milan, Italy (1 x weekday); Utrecht, The Netherlands (1 x weekday, 2 x weekend); Zurich, Switzerland (1 x weekday, 2 x weekend). As shown in Table 3 of Scientific Article 7, only MDPV, mephedrone and methylone were found. The highest concentration of any compound was for mephedrone, with a concentration of 106ng/L found

in South West UK, showing its popularity is far higher in the United Kingdom than in other cities of continental Europe, although it was also quantified in Copenhagen and detected in Oslo and Brussels, although at concentrations lower than LOQ. MDPV was found in Milan, Oslo and Zurich at low concentrations (2-6ng/L), while methylone was found in Utrecht and Copenhagen (5-12ng/L) and detected in Bristol at a concentration lower than LOQ.

CHAPTER 5

REFERENCES

- Abraham, M.H., Ibrahim, A., Zissimos, A.M., 2004. Determination of sets of solute descriptors from chromatographic measurements. *J. Chromatogr. A* 1037, 29–47. doi:10.1016/j.chroma.2003.12.004
- Agüera, A., Martínez Bueno, M.J., Fernández-Alba, A.R., 2013. New trends in the analytical determination of emerging contaminants and their transformation products in environmental waters. *Environ. Sci. Pollut. Res. Int.* 20, 3496–515. doi:10.1007/s11356-013-1586-0
- Anumol, T., Snyder, S. a., 2015. Rapid analysis of trace organic compounds in water by automated online solid-phase extraction coupled to liquid chromatography–tandem mass spectrometry. *Talanta* 132, 77–86. doi:10.1016/j.talanta.2014.08.011
- Backe, W.J., Field, J. a., 2012. Is SPE necessary for environmental analysis? A quantitative comparison of matrix effects from large-volume injection and solid-phase extraction based methods. *Environ. Sci. Technol.* 46, 6750–6758. doi:10.1021/es300235z
- Banta-Green, C.J., Field, J.A., Chiaia, A.C., Sudakin, D.L., Power, L., De Montigny, L., 2009. The spatial epidemiology of cocaine, methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) use: A demonstration using a population measure of community drug load derived from municipal wastewater. *Addiction* 104, 1874–1880. doi:10.1111/j.1360-0443.2009.02678.x
- Barron, L., Havel, J., Purcell, M., Szpak, M., Kelleher, B., Paull, B., 2009. Predicting sorption of pharmaceuticals and personal care products onto soil and digested sludge using artificial neural networks. *Analyst* 134, 663–70. doi:10.1039/b817822d
- Barron, L.P., McEneff, G.L., 2016. Gradient liquid chromatographic retention time prediction for suspect screening applications: A critical assessment of a generalised artificial neural network-based approach across 10 multi-residue reversed-phase analytical methods. *Talanta* 147, 261–270. doi:10.1016/j.talanta.2015.09.065
- Bäuerlein, P.S., ter Laak, T.L., Hofman-Caris, R.C.H.M., de Voogt, P., Droge, S.T.J., 2012. Removal of charged micropollutants from water by ion-exchange polymers - Effects of competing electrolytes. *Water Res.* 46, 5009–5018. doi:10.1016/j.watres.2012.06.048
- Bijlsma, L., Emke, E., Hernández, F., de Voogt, P., Hernandez, F., 2013. Performance of the linear ion trap Orbitrap mass analyzer for qualitative and quantitative analysis of drugs of abuse and relevant metabolites in sewage water. *Anal. Chim. Acta* 768, 102–110. doi:10.1016/j.aca.2013.01.010

- Bijlsma, L., Beltrán, E., Boix, C., Sancho, J. V., Hernández, F., 2014. Improvements in analytical methodology for the determination of frequently consumed illicit drugs in urban wastewater. *Anal. Bioanal. Chem.* 406, 4261–72. doi:10.1007/s00216-014-7818-4
- Bletsou, A.A., Jeon, J., Hollender, J., Archontaki, E., Thomaidis, N.S., 2015. Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment. *Trends Anal. Chem.* 66, 32–44.
- Borova, V.L., Gago-Ferrero, P., Pistos, C., Thomaidis, N.S., 2015. Multi-residue determination of 10 selected new psychoactive substances in wastewater samples by liquid chromatography–tandem mass spectrometry. *Talanta* 144, 592–603. doi:10.1016/j.talanta.2015.06.080
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324
- Burgard, D.A., Banta-Green, C., Field, J.A., 2014. Working upstream: how far can you go with sewage-based drug epidemiology? *Environ. Sci. Technol.* 48, 1362–8. doi:10.1021/es4044648
- Buszewski, B., Noga, S., 2012. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Anal. Bioanal. Chem.* 402, 231–247. doi:10.1007/s00216-011-5308-5
- Caspar, A.T., Helfer, A.G., Michely, J.A., Auwarter, V., Brandt, S.D., Meyer, M.R., Maurer, H.H., 2015. Studies on the metabolism and toxicological detection of the new psychoactive designer drug 2-(4-iodo-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl]ethanamine (25I-NBOMe) in human and rat urine using GC-MS, LC-MS, and LC-HR-MS/MS. *Anal. Bioanal. Chem.* 6697–6719. doi:10.1007/s00216-015-8828-6
- Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Hernández, F., Reid, M., Ort, C., Thomas, K. V., van Nuijs, A.L.N., de Voogt, P., Zuccato, E., 2013. Evaluation of uncertainties associated with the determination of community drug use through the measurement of sewage drug biomarkers. *Environ. Sci. Technol.* 47, 1452–60. doi:10.1021/es302722f
- Castiglioni, S., Bijlsma, L., Covaci, A., Emke, E., Harman, C., Hernández, F., Kasprzyk-Hordern, B., Ort, C., van Nuijs, A.L.N., de Voogt, P., Zuccato, E., 2016. Estimating community drug use through wastewater-based epidemiology, in: Castiglioni, S. (Ed.), *Assessing Illicit Drugs in Wastewater: Advances in Wastewater-Based Drug Epidemiology*. pp. 17–33.

- Castiglioni, S., Gracia-Lor, E., 2016. Target drug residues in wastewater, in: Castiglioni, S. (Ed.), *Assessing Illicit Drugs in Wastewater: Advances in Wastewater-Based Drug Epidemiology*. pp. 35–43.
- Chimalakonda, K.C., Seely, K.A., Bratton, S.M., Brents, L.K., Moran, C.L., Endres, G.W., James, L.P., Hollenberg, P.F., Prather, P.L., Radomska-Pandya, A., Moran, J.H., 2012. Cytochrome P450-Mediated Oxidative Metabolism of Abused Synthetic Cannabinoids Found in K2/Spice: Identification of Novel Cannabinoid Receptor Ligands. *Drug Metab. Dispos.* 40, 2174–2184. doi:10.1124/dmd.112.047530
- Croley, T.R., White, K.D., Callahan, J.H., Musser, S.M., 2012. The chromatographic role in high resolution mass spectrometry for non-targeted analysis. *J. Am. Soc. Mass Spectrom.* 23, 1569–78. doi:10.1007/s13361-012-0392-0
- Dalby, A., Nourse, J.G., Hounshell, W.D., Gushurst, A.K.I., Grier, D.L., Leland, B. a, Laufer, J., 1992. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Model.* 32, 244–255. doi:10.1021/ci00007a012
- Daughton, C.G., 2014. The Matthew Effect and widely prescribed pharmaceuticals lacking environmental monitoring: Case study of an exposure-assessment vulnerability. *Sci. Total Environ.* 466-467, 315–325. doi:10.1016/j.scitotenv.2013.06.111
- Daughton, C.G., 2012. Using biomarkers in sewage to monitor community-wide human health: isoprostanes as conceptual prototype. *Sci. Total Environ.* 424, 16–38. doi:10.1016/j.scitotenv.2012.02.038
- Daughton, C.G., 2001. *Illicit Drugs in Municipal Sewage*. pp. 348–364. doi:10.1021/bk-2001-0791.ch020
- de Voogt, P., Zurano, L., Serné, P., Haftka, J.J.H., 2012. Experimental hydrophobicity parameters of perfluorinated alkylated substances from reversed-phase high-performance liquid chromatography. *Environ. Chem.* 9, 564–570. doi:10.1071/EN12132
- Du, P., Li, K., Li, J., Xu, Z., Fu, X., Yang, J., Zhang, H., Li, X., 2015. Methamphetamine and ketamine use in major Chinese cities, a nationwide reconnaissance through sewage-based epidemiology. *Water Res.* 84, 76–84. doi:10.1016/j.watres.2015.07.025

- EMCDDA, 2015. European Drug Report: Trends and Developments.
- EMCDDA, 2009. Understanding the Spice Phenomenon. *Themat. Pap.* 25. doi:10.2810/27063
- Emke, E., Evans, S., Kasprzyk-Hordern, B., de Voogt, P., 2014. Enantiomer profiling of high loads of amphetamine and MDMA in communal sewage: a Dutch perspective. *Sci. Total Environ.* 487, 666–72. doi:10.1016/j.scitotenv.2013.11.043
- Farré, M., Holčápek, M., Schoenmakers, P.J., Kantiani, L., Petrovic, M., Pérez, S., Barceló, D., 2012. Achievements and future trends in the analysis of emerging organic contaminants in environmental samples by mass spectrometry and bioanalytical techniques. *J. Chromatogr. A* 1259, 86–99.
- Fatemi, M.H., Ghorbanzad'e, M., Baher, E., 2010. Quantitative Structure Retention Relationship Modeling of Retention Time for Some Organic Pollutants. *Anal. Lett.* 43, 823–835. doi:10.1080/00032710903486294
- Ferrer, I., Thurman, E.M., 2013. Analysis of Pharmaceuticals in Drinking Water, Groundwater, Surface Water, and Wastewater, in: Petrovic, M., Barcelo, D., Perez, S. (Eds.), *Analysis, Removal, Effects and Risk of Pharmaceuticals in the Water Cycle — Occurrence and Transformation in the Environment*. pp. 91–128. doi:10.1016/B978-0-444-62657-8.00004-5
- Fornal, E., 2013. Identification of substituted cathinones: 3,4-Methylenedioxy derivatives by high performance liquid chromatography-quadrupole time of flight mass spectrometry. *J. Pharm. Biomed. Anal.* 81-82, 13–19. doi:10.1016/j.jpba.2013.03.016
- Gago-Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S., 2015. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* 49, 12333–12341. doi:10.1021/acs.est.5b03454
- Gerlich, M., Neumann, S., 2013. MetFusion: integration of compound identification strategies. *J. Mass Spectrom.* 48, 291–8. doi:10.1002/jms.3123
- German, C.L., Fleckenstein, A.E., Hanson, G.R., 2014. Bath salts and synthetic cathinones: An emerging designer drug phenomenon. *Life Sci.* 97, 2–8. doi:10.1016/j.lfs.2013.07.023

- Giaginis, C., Tsantili-Kakoulidou, A., 2012. Quantitative Structure–Retention Relationships as Useful Tool to Characterize Chromatographic Systems and Their Potential to Simulate Biological Processes. *Chromatographia* 76, 211–226. doi:10.1007/s10337-012-2374-6
- Gibbons, S., 2012. “Legal highs”--novel and emerging psychoactive drugs: a chemical overview for the toxicologist. *Clin. Toxicol. (Phila)*. 50, 15–24. doi:10.3109/15563650.2011.645952
- Gonzalez-Mariño, I., Quintana, J.B., Rodriguez, I., Gonzalez-Diez, M., Cela, R., 2012. Screening and Selective Quantification of Illicit Drugs in Wastewater by Mixed-Mode Solid-Phase Extraction and Quadrupole-Time-of-Flight Liquid Chromatography–Mass Spectrometry. *Anal. Chem.* 84, 1708–1717. doi:10.1021/ac202989e
- Goudarzi, N., Shahsavani, D., 2012. Application of a random forests (RF) method as a new approach for variable selection and modelling in a QSRR study to predict the relative retention time of some polybrominated diphenylethers (PBDEs). *Anal. Methods* 4, 3733. doi:10.1039/c2ay25484k
- Grabenaus, M., Krol, W.L., Wiley, J.L., Thomas, B.F., 2012. Analysis of Synthetic Cannabinoids Using High-Resolution Mass Spectrometry and Mass Defect Filtering: Implications for Nontargeted Screening of Designer Drugs. *Anal. Chem.* 84, 5574–5581. doi:10.1021/ac300509h
- Guillarme, D., Schappler, J., Rudaz, S., Veuthey, J.-L., 2010. Coupling ultra-high-pressure liquid chromatography with mass spectrometry. *TrAC Trends Anal. Chem.* 29, 15–27. doi:10.1016/j.trac.2009.09.008
- Hancock, T., Put, R., Coomans, D., Vander Heyden, Y., Everingham, Y., 2005. A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies. *Chemom. Intell. Lab. Syst.* 76, 185–196. doi:10.1016/j.chemolab.2004.11.001
- Héberger, K., 2007. Quantitative structure-(chromatographic) retention relationships. *J. Chromatogr. A* 1158, 273–305. doi:10.1016/j.chroma.2007.03.108
- Hernández, F., Bijlsma, L., Sancho, J. V., Díaz, R., Ibáñez, M., 2011. Rapid wide-scope screening of drugs of abuse, prescription drugs with potential for abuse and their metabolites in influent and effluent urban wastewater by ultrahigh pressure liquid chromatography-quadrupole-time-of-flight-mass spectrometry. *Anal. Chim. Acta* 684, 87–97. doi:10.1016/j.aca.2010.10.043

- Hernández, F., Ibáñez, M., Portolés, T., Cervera, M.I., Sancho, J. V, López, F.J., 2015. Advancing towards universal screening for organic pollutants in waters. *J. Hazard. Mater.* 282, 86–95. doi:10.1016/j.jhazmat.2014.08.006
- Hernández, F., Pozo, Ó.J., Sancho, J. V., López, F.J., Marín, J.M., Ibáñez, M., 2005. Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC–MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. *TrAC Trends Anal. Chem.* 24, 596–612. doi:10.1016/j.trac.2005.04.007
- Hernández, F., Pozo, Ó.J., Sancho, J. V., López, F.J., Marín, J.M., Ibáñez, M., 2005. Strategies for quantification and confirmation of multi-class polar pesticides and transformation products in water by LC–MS2 using triple quadrupole and hybrid quadrupole time-of-flight analyzers. *TrAC Trends Anal. Chem.* 24, 596–612. doi:10.1016/j.trac.2005.04.007
- Hill, S.L., Thomas, S.H.L., 2016. Drugs of abuse. *Medicine (Baltimore)*. 1–10. doi:10.1016/j.mpmed.2015.12.030
- Hug, C., Ulrich, N., Schulze, T., Brack, W., Krauss, M., 2014. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* 184, 25–32. doi:10.1016/j.envpol.2013.07.048
- Ibáñez, M., Bijlsma, L., van Nuijs, A.L.N., Sancho, J. V, Haro, G., Covaci, A., Hernández, F., 2013. Quadrupole-time-of-flight mass spectrometry screening for synthetic cannabinoids in herbal blends. *J. Mass Spectrom.* 48, 685–94. doi:10.1002/jms.3217
- Ibáñez, M., Pozo, Ó.J., Sancho, J. V., Orengo, T., Haro, G., Hernández, F., 2016. Analytical strategy to investigate 3,4-methylenedioxypyrovalerone (MDPV) metabolites in consumers' urine by high-resolution mass spectrometry. *Anal. Bioanal. Chem.* 408, 151–164. doi:10.1007/s00216-015-9088-1
- Ibáñez, M., Sancho, J. V., Bijlsma, L., van Nuijs, A.L.N., Covaci, A., Hernández, F., 2014. Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances. *TrAC Trends Anal. Chem.* 57, 107–117. doi:10.1016/j.trac.2014.02.009
- Ibáñez, M., Borova, V., Boix, C., Aalizadeh, R., Bade, R., Thomaidis, N.S., Hernández, F., 2016. UHPLC-QTOF MS screening of pharmaceuticals and their metabolites in treated wastewater samples from Athens. *J. Hazard. Mater.* doi:10.1016/j.jhazmat.2016.03.078

- Irvine, R.J., Kostakis, C., Felgate, P.D., Jaehne, E.J., Chen, C., White, J.M., 2011. Population drug use in Australia: A wastewater analysis. *Forensic Sci. Int.* 210, 69–73. doi:10.1016/j.forsciint.2011.01.037
- Ivanciuc, O., 2007. Applications of support vector machines in chemistry. *Rev. Comput. Chem.* 23, 291–400. doi:10.1002/9780470116449.ch6
- Kaliszan, R., 1992. Quantitative structure-retention relationships. *Anal. Chem.* 64, 619A–631A. doi:10.1021/ac00035a001
- Kaliszan, R., 1977. Correlation between the retention indices and the connectivity indices of alcohols and methyl esters with complex cyclic structure. *Chromatographia* 10, 529–531. doi:10.1007/BF02262911
- Kaliszan, R., Baczek, T., Bucinkński, A., Buszewski, B., Sztupecka, M., 2003. Prediction of gradient retention from the linear solvent strength (LSS) model, quantitative structure-retention relationships (QSRR), and artificial neural networks (ANN). *J. Sep. Sci.* 26, 271–282. doi:10.1002/jssc.200390033
- Kamata, H.T., Shima, N., Zaitso, K., Kamata, T., Miki, a, Nishikawa, M., Katagi, M., Tsuchihashi, H., 2006. Metabolism of the recently encountered designer drug, methylone, in humans and rats. *Xenobiotica.* 36, 709–23. doi:10.1080/00498250600780191
- Kankaanpää, A., Ariniemi, K., Heinonen, M., Kuoppasalmi, K., Gunnar, T., 2014. Use of illicit stimulant drugs in Finland: A wastewater study in ten major cities. *Sci. Total Environ.* 487, 696–702. doi:10.1016/j.scitotenv.2013.11.095
- Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., Hollender, J., 2009. Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. *Environ. Sci. Technol.* 43, 7039–7046. doi:10.1021/es901979h
- Khan, U., van Nuijs, a L.N., Li, J., Maho, W., Du, P., Li, K.Y., Hou, L.L., Zhang, J.Y., Meng, X.Z., Li, X.Q., Covaci, a, 2014. Application of a sewage-based approach to assess the use of ten illicit drugs in four Chinese megacities. *Sci. Total Environ.* 487, 710–721. doi:10.1016/j.scitotenv.2014.01.043

- Kim, K.Y., Lai, F.Y., Kim, H.-Y., Thai, P.K., Mueller, J.F., Oh, J.-E., 2015. The first application of wastewater-based drug epidemiology in five South Korean cities. *Sci. Total Environ.* 524-525, 440–6. doi:10.1016/j.scitotenv.2015.04.065
- Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 8, 105. doi:10.1186/1471-2105-8-105
- Kinyua, J., Covaci, A., Maho, W., McCall, A.-K., Neels, H., van Nuijs, A.L.N., 2015. Sewage-based epidemiology in monitoring the use of new psychoactive substances: Validation and application of an analytical method using LC-MS/MS. *Drug Test. Anal.* 7, 812–818. doi:10.1002/dta.1777
- Kinyua, J., Negreira, N., Ibáñez, M., Bijlsma, L., Hernández, F., Covaci, A., van Nuijs, A.L.N., 2015b. A data-independent acquisition workflow for qualitative screening of new psychoactive substances in biological samples. *Anal. Bioanal. Chem.* 407, 8773–8785. doi:10.1007/s00216-015-9036-0
- Krauss, M., Singer, H., Hollender, J., 2010. LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* 397, 943–951. doi:10.1007/s00216-010-3608-9
- Lai, F.Y., Erratico, C., Kinyua, J., Mueller, J.F., Covaci, A., van Nuijs, A.L.N., 2015. Liquid chromatography-quadrupole time-of-flight mass spectrometry for screening in vitro drug metabolites in humans: investigation on seven phenethylamine-based designer drugs. *J. Pharm. Biomed. Anal.* 114, 355–375. doi:10.1016/j.jpba.2015.06.016
- Lai, F.Y., Bruno, R., Hall, W., Gartner, C., Ort, C., Kirkbride, P., Prichard, J., Thai, P.K., Carter, S., Mueller, J.F., 2013a. Profiles of illicit drug use during annual key holiday and control periods in Australia: wastewater analysis in an urban, a semi-rural and a vacation area. *Addiction* 108, 556–65. doi:10.1111/add.12006
- Lai, F.Y., Bruno, R., Leung, H.W., Thai, P.K., Ort, C., Carter, S., Thompson, K., Lam, P.K.S., Mueller, J.F., 2013b. Estimating daily and diurnal variations of illicit drug use in Hong Kong: A pilot study of using wastewater analysis in an Asian metropolitan city. *Forensic Sci. Int.* 233, 126–132. doi:10.1016/j.forsciint.2013.09.003
- Leendert, V., Van Langenhove, H., Demeestere, K., 2015. Trends in liquid chromatography coupled to high-resolution mass spectrometry for multi-residue analysis of organic micropollutants in aquatic environments. *TrAC Trends Anal. Chem.* 67, 192–208. doi:10.1016/j.trac.2015.01.010

- Leffler, A.M., Smith, P.B., de Armas, A., Dorman, F.L., 2014. The analytical investigation of synthetic street drugs containing cathinone analogs. *Forensic Sci. Int.* 234, 50–56. doi:10.1016/j.forsciint.2013.08.021
- Little, J.L., Williams, A.J., Pshenichnov, A., Tkachenko, V., 2012. Identification of “Known Unknowns” Utilizing Accurate Mass Data and ChemSpider. *J. Am. Soc. Mass Spectrom.* 23, 179–185. doi:10.1007/s13361-011-0265-y
- Livingstone, D., 2003. Theoretical Property Predictions. *Curr. Top. Med. Chem.* 3, 1171–1192. doi:10.2174/1568026033452078
- Mardal, M., Meyer, M.R., 2014. Studies on the microbial biotransformation of the novel psychoactive substance methylenedioxypropylamphetamine (MDPV) in wastewater by means of liquid chromatography-high resolution mass spectrometry/mass spectrometry. *Sci. Total Environ.* 493, 588–595. doi:10.1016/j.scitotenv.2014.06.016
- Martínez Bueno, M.J., Ulaszewska, M.M., Gomez, M.J., Hernando, M.D., Fernández-Alba, A.R., 2012. Simultaneous measurement in mass and mass/mass mode for accurate qualitative and quantitative screening analysis of pharmaceuticals in river water. *J. Chromatogr. A* 1256, 80–8. doi:10.1016/j.chroma.2012.07.038
- Matuszewski, B.K., Constanzer, M.L., Chavez-Eng, C.M., 2003. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal. Chem.* 75, 3019–3030. doi:10.1021/ac020361s
- Maurer, H.H., 2010. Chemistry, pharmacology, and metabolism of emerging drugs of abuse. *Ther. Drug Monit.* 32, 544–9. doi:10.1097/FTD.0b013e3181eea318
- McCall, A.-K., Bade, R., Kinyua, J., Lai, F.Y., Thai, P.K., Covaci, A., Bijlsma, L., van Nuijs, A.L.N., Ort, C., 2016. Critical review on the stability of illicit drugs in sewers and wastewater samples. *Water Res.* 88, 933–947. doi:10.1016/j.watres.2015.10.040
- Metcalf, C., Tindale, K., Li, H., Rodayan, A., Yargeau, V., 2010. Illicit drugs in Canadian municipal wastewater and estimates of community drug use. *Environ. Pollut.* 158, 3179–3185. doi:10.1016/j.envpol.2010.07.002
- Meyer, M.R., Du, P., Schuster, F., Maurer, H.H., 2010. Studies on the metabolism of the α -pyrrolidinophenone designer drug methylenedioxy-propylamphetamine (MDPV) in rat and human

- urine and human liver microsomes using GC-MS and LC-high-resolution MS and its detectability in urine by GC-MS. *J. Mass Spectrom.* 45, 1426–1442. doi:10.1002/jms.1859
- Miller, T.H., Musenga, A., Cowan, D.A., Barron, L.P., 2013. Prediction of chromatographic retention time in high-resolution anti-doping screening data using artificial neural networks. *Anal. Chem.* 85, 10330–7. doi:10.1021/ac4024878
- Moran, C.L., Le, V.H., Chimalakonda, K.C., Smedley, A.L., Lackey, F.D., Owen, S.N., Kennedy, P.D., Endres, G.W., Ciske, F.L., Kramer, J.B., Kornilov, A.M., Bratton, L.D., Dobrowolski, P.J., Wessinger, W.D., Fantegrossi, W.E., Prather, P.L., James, L.P., Radominska-Pandya, A., Moran, J.H., 2011. Quantitative measurement of JWH-018 and JWH-073 metabolites excreted in human urine. *Anal. Chem.* 83, 4228–4236. doi:10.1021/ac2005636
- Müller, A., Schulz, W., Ruck, W.K.L., Weber, W.H., 2011. A new approach to data evaluation in the non-target screening of organic trace substances in water analysis. *Chemosphere* 85, 1211–1219. doi:10.1016/j.chemosphere.2011.07.009
- National Alliance for Model State Drug Laws, 2012. Cathinones Derivates Trade Name and Chemical Compound Chart.
- Niessen, W.M.A., Manini, P., Andreoli, R., 2006. Matrix effects in quantitative pesticide analysis using liquid chromatography–mass spectrometry. *Mass Spectrom. Rev.* 25, 881–899. doi:10.1002/mas.20097
- Nurmi, J., Pellinen, J., Rantalainen, A.-L., 2012. Critical evaluation of screening techniques for emerging environmental contaminants based on accurate mass measurements with time-of-flight mass spectrometry. *J. Mass Spectrom.* 47, 303–12. doi:10.1002/jms.2964
- Nurmi, J., Pellinen, J., 2011. Multiresidue method for the analysis of emerging contaminants in wastewater by ultra performance liquid chromatography-time-of-flight mass spectrometry. *J. Chromatogr. A* 1218, 6712–9. doi:10.1016/j.chroma.2011.07.071
- OECD, 2004. OECD guidelines for testing of chemicals. 117. Partition coefficient (n-octanol/water), High Performance Liquid Chromatography (HPLC) method. Extrapolation 1–11. doi:10.1787/9789264069824-en
- Ort, C., van Nuijs, A.L.N., Berset, J.-D., Bijlsma, L., Castiglioni, S., Covaci, A., de Voogt, P., Emke, E., Fatta-Kassinos, D., Griffiths, P., Hernández, F., González-Mariño, I., Grabic, R., Kasprzyk-Hordern, B., Mastroianni, N., Meierjohann, A., Nefau, T., Ostman, M., Pico, Y., Racamonde, I., Reid, M., Slobodnik, J., Terzic, S., Thomaidis, N., Thomas, K. V., 2014. Spatial differences and

- temporal changes in illicit drug use in Europe quantified by wastewater analysis. *Addiction* 109, 1338–52. doi:10.1111/add.12570
- Plumb, R.S., Johnson, K.A., Rainville, P., Smith, B.W., Wilson, I.D., Castro-Perez, J.M., Nicholson, J.K., 2006. UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* 20, 1989–1994. doi:10.1002/rcm.2550
- Poklis, J.L., Nanco, C.R., Troendle, M.M., Wolf, C.E., Poklis, A., 2014. Determination of 4-bromo-2,5-dimethoxy-N-[(2-methoxyphenyl)methyl]-benzeneethanamine (25B-NBOMe) in serum and urine by high performance liquid chromatography with tandem mass spectrometry in a case of severe intoxication. *Drug Test Anal* 6, 764–9. doi:10.1002/dta.1522.Determination
- Pourbasheer, E., Aalizadeh, R., Ardabili, J.S., Ganjali, M.R., 2015. QSPR study on solubility of some fullerenes derivatives using the genetic algorithms — Multiple linear regression. *J. Mol. Liq.* 204, 162–169. doi:10.1016/j.molliq.2015.01.028
- Reid, M.J., Langford, K.H., Morland, J., Thomas, K. V., 2011. Analysis and interpretation of specific ethanol metabolites, ethyl sulfate, and ethyl glucuronide in sewage effluent for the quantitative measurement of regional alcohol consumption. *Alcohol Clin Exp Res* 35, 1593–1599. doi:10.1111/j.1530-0277.2011.01505.x
- Reid, M.J., Derry, L., Thomas, K. V., 2014. Analysis of new classes of recreational drugs in sewage: Synthetic cannabinoids and amphetamine-like substances. *Drug Test. Anal.* 6, 72–79. doi:10.1002/dta.1461
- Richardson, S.D., Ternes, T.A., 2014. Water analysis: emerging contaminants and current issues. *Anal. Chem.* 86, 2813–48. doi:10.1021/ac500508t
- Ryu, Y., Reid, M.J., Thomas, K. V., 2015. Liquid chromatography–high resolution mass spectrometry with immunoaffinity clean-up for the determination of the oxidative stress biomarker 8-iso-prostaglandin F₂α in wastewater. *J. Chromatogr. A* 1409, 146–151. doi:10.1016/j.chroma.2015.07.060
- Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J., 2015. Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. *Anal. Chem.* acs.analchem.5b02905. doi:10.1021/acs.analchem.5b02905

- Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J., 2014. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. *Environ. Sci. Technol.* 48, 1811–8. doi:10.1021/es4044374
- Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* 48, 2097–2098. doi:10.1021/es5002105
- Senta, I., Gracia-Lor, E., Borsotti, A., Zuccato, E., Castiglioni, S., 2015. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. *Water Res.* 74, 23–33. doi:10.1016/j.watres.2015.02.002
- Senta, I., Krizman, I., Ahel, M., Terzic, S., 2015. Multiresidual analysis of emerging amphetamine-like psychoactive substances in wastewater and river water. *J. Chromatogr. A* 1425, 204–212. doi:10.1016/j.chroma.2015.11.043
- Shen, Q., Lü, Q.-Z., Jiang, J.-H., Shen, G.-L., Yu, R.-Q., 2003. Quantitative structure-activity relationships (QSAR): studies of inhibitors of tyrosine kinase. *Eur. J. Pharm. Sci.* 20, 63–71. doi:10.1016/S0928-0987(03)00170-2
- Shinoda, K., Sugimoto, M., Tomita, M., Ishihama, Y., 2008. Informatics for peptide retention properties in proteomic LC-MS. *Proteomics* 8, 787–798. doi:10.1002/pmic.200700692
- Stahnke, H., Kittlaus, S., Kempe, G., Alder, L., 2012. Reduction of matrix effects in liquid chromatography-electrospray ionization-mass spectrometry by dilution of the sample extracts: how much dilution is needed? *Anal. Chem.* 84, 1474–82. doi:10.1021/ac202661j
- Stanstrup, J., Neumann, S., Vrhovšek, U., 2015. PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems. *Anal. Chem.* 87, 9421–9428. doi:10.1021/acs.analchem.5b02287
- Subedi, B., Kannan, K., 2014. Mass loading and removal of select illicit drugs in two wastewater treatment plants in New York State and estimation of illicit drug usage in communities through wastewater analysis. *Environ. Sci. Technol.* 48, 6661–6670. doi:10.1021/es501709a
- Talete S.R.L., 2013. Dragon 6.
- Tang, M.H., Ching, C.K., Tsui, M.S., Chu, F.K., Mak, T.W., 2014. Two cases of severe intoxication associated with analytically confirmed use of the novel psychoactive substances 25B-NBOMe

- and 25C-NBOMe. *Clin Toxicol* 52, 561–565. doi:10.3109/15563650.2014.909932
- Taylor, P.J., 2005. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clin. Biochem.* 38, 328–334. doi:10.1016/j.clinbiochem.2004.11.007
- Tetko, I. V, Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V., Radchenko, E., Zefirov, N.S., Makarenko, A.S., Tanchuk, V.Y., Prokopenko, V. V, 2005. Virtual computational chemistry laboratory - design and description. *J. Comput. Aided. Mol. Des.* 19, 453–463. doi:10.1007/s10822-005-8694-y
- Thomas, K. V, Bijlsma, L., Castiglioni, S., Covaci, A., Emke, E., Grabic, R., Hernández, F., Karolak, S., Kasprzyk-Hordern, B., Lindberg, R.H., Lopez de Alda, M., Meierjohann, A., Ort, C., Pico, Y., Quintana, J.B.J.B., Reid, M., Rieckermann, J.J., Terzic, S., van Nuijs, A.L.N., de Voogt, P., Hernandez, F., 2012. Comparing illicit drug use in 19 European cities through sewage analysis. *Sci. Total Environ.* 432, 432–9. doi:10.1016/j.scitotenv.2012.06.069
- Thomas, K. V, Reid, M.J., 2011. What Else Can the Analysis of Sewage for Urinary Biomarkers Reveal About Communities? *Environ. Sci. Technol.* 45, 7611–7612. doi:10.1021/es202522d
- Ulrich, N., Schüürmann, G., Brack, W., 2011. Linear Solvation Energy Relationships as classifiers in non-target analysis—A capillary liquid chromatography approach. *J. Chromatogr. A* 1218, 8192–8196. doi:10.1016/j.chroma.2011.09.031
- UNODC, 2014. Global Synthetic Drugs Assessment: Amphetamine-type stimulants and new psychoactive substances.
- Uralets, V., Rana, S., Morgan, S., Ross, W., 2014. Testing for designer stimulants: Metabolic profiles of 16 synthetic cathinones excreted free in human urine. *J. Anal. Toxicol.* 38, 233–241. doi:10.1093/jat/bku021
- van Nuijs, A.L.N., Castiglioni, S., Tarcomnicu, I., Postigo, C., Lopez de Alda, M., Neels, H., Zuccato, E., Barcelo, D., Covaci, A., 2011. Illicit drug consumption estimations derived from wastewater analysis: a critical review. *Sci. Total Environ.* 409, 3564–77. doi:10.1016/j.scitotenv.2010.05.030
- van Nuijs, A.L.N., Tarcomnicu, I., Covaci, A., 2011. Application of hydrophilic interaction chromatography for the analysis of polar contaminants in food and environmental samples. *J. Chromatogr. A* 1218, 5964–5974. doi:10.1016/j.chroma.2011.01.075

- van Nuijs, A.L.N., Gheorghe, A., Jorens, P.G., Maudens, K., Neels, H., Covaci, A., 2013. Optimization, validation, and the application of liquid chromatography-tandem mass spectrometry for the analysis of new drugs of abuse in wastewater. *Drug Test. Anal.* 861–867. doi:10.1002/dta.1460
- Vazquez-Roig, P., Blasco, C., Picó, Y., 2013. Advances in the analysis of legal and illegal drugs in the aquatic environment. *Trends Anal. Chem.* 50, 65–77. doi:10.1016/j.trac.2013.04.008
- Venhuis, B.J., de Voogt, P., Emke, E., Causanilles, A., Keizers, P.H.J., 2014. Success of rogue online pharmacies: sewage study of sildenafil in the Netherlands. *BMJ* 349, 4317–4317. doi:10.1136/bmj.g4317
- Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* 28, 31–36. doi:10.1021/ci00057a005
- Zancajo, V.M.R., Brito, J., Carrasco, M.P., Bronze, M.R., Moreira, R., Lopes, A., 2014. Analytical profiles of “legal highs” containing cathinones available in the area of Lisbon, Portugal. *Forensic Sci. Int.* 244, 102–110. doi:10.1016/j.forsciint.2014.08.010
- Zuba, D., 2012. Identification of cathinones and other active components of “legal highs” by mass spectrometric methods. *TrAC - Trends Anal. Chem.* 32, 15–30. doi:10.1016/j.trac.2011.09.009
- Zuccato, E., Chiabrando, C., Castiglioni, S., Calamari, D., Bagnati, R., Schiarea, S., Fanelli, R., 2005. Cocaine in surface waters: a new evidence-based tool to monitor community drug abuse. *Environ. Heal. A Glob. Access Sci. Source* 4, 14. doi:10.1186/1476-069X-4-14

CHAPTER 6

CONCLUSIONS

CONCLUSIONES

6.1 Conclusions

The general conclusion of this thesis is that advanced analytical techniques, based on liquid chromatography coupled to low and high-resolution mass spectrometry, and incorporating retention time prediction and target, suspect and non-target analyses are of great utility for the detection and identification of emerging contaminants in water. The use of (hyphenated) high resolution mass spectrometry such as triple quadrupole (QqQ), quadrupole time-of-flight (QTOF) and Linear ion Trap (LTQ) Orbitrap, allows comprehensive analysis of emerging contaminants in environmental waters, through the use of target (QqQ), suspect (QTOF and LTQ-Orbitrap) and non-target (LTQ-Orbitrap) analyses. Furthermore, the complementary use of retention time prediction made the detection of compounds (especially for which standards were unavailable) much less tedious and enabled higher confidence in their tentative identification.

From the work performed in this thesis, several specific conclusions can be garnered:

1. High resolution mass spectrometry, especially QTOF and LTQ-Orbitrap are very powerful instruments for the qualitative suspect and non-target screening of emerging contaminants in aquatic samples.
2. QTOF in MS^E mode facilitates the detection and tentative identification of pharmaceuticals, illicit drugs (PIDs) and their metabolites, with high energy (corresponding to fragment ions) and low energy (corresponding primarily to the parent molecule) information gained in a single injection.
3. It has been shown that there is a geographic difference in the detection of PIDs, with samples from Colombia, Italy and Spain having been analysed. While the samples from Italy and Spain were quite similar - Irbesartan (antihypertensive), valsartan (angiotensin II antagonist) and two antibiotics ofloxacin and sulfamethoxazole were the most commonly identified compounds – in the Colombian samples the beta-blocker metoprolol, the lipid regulator gemfibrozil, and the X-ray agent iopromide as well as metabolites of losartan and carbamazepine were found.
4. The use of non-target analyses in a hidden target approach provides a lot of information. Furthermore, even without reference standards, numerous compounds can be tentatively identified.
5. Retention time prediction has been shown as a great complementary technique in the detection and identification of PIDs and their metabolites.
6. A retention time prediction approach based solely on log K_{ow} is not only cheap and simple, but also led to between 70-80% of all compounds having a predicted retention time within 2 minutes of their experimental.
7. A more complex retention time prediction approach, based on artificial neural networks, showed a greater predictive capacity, with up to 95% of all compounds having a predicted retention time within 2 minutes of their experimental. This allowed greater confidence in the tentative identification of compounds for which standards were unavailable.

8. Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) with triple quadrupole mass analyser is a powerful analytical technique for the quantification of new psychoactive substances (NPS) in wastewater. The robustness, sensitivity and selectivity of this technique make it highly appropriate for the determination of these compounds at the low concentrations commonly found in complex-matrix samples such as urban wastewater.
9. Additional studies of filtration are necessary to account for possible losses during sample treatment. Vacuum filtration (using membrane filters) was found to have under 5% recovery for the three NBOMe compounds as well as naphyrone. Syringe filters (of PTFE, polypropylene and nylon) were found to have better recoveries for all compounds, with the Nylon 13mm, 0.45µm proving to be the best, which had recoveries of 98-105% for all compounds
10. Mobile phase additives, sample injection volume and desolvation and cone gas flow rates all contribute to optimising chromatographic peak shape and mass spectrometric optimisation.
11. Stability is an important parameter for WBE of NPS, with acidification (to pH 2) necessary to reduce transformation and degradation for all NPS investigated at 20°C, 4°C and -20°C for up to seven days. If left unacidified, samples should be analysed within 48 hours (if stored at -20°C or 4°C), while at 20°C, N-ethylcathinone and mephedrone showed low stability within 24 hours.
12. The acquisition of two or three transitions allows the simultaneous quantification and confirmation of various NPS.

6.2 Conclusiones

La conclusión general de esta tesis es que las técnicas analíticas avanzadas, basadas en la cromatografía líquida acoplada a espectrometría de masas de baja y de alta resolución, para el análisis “target”, “suspect” y “non-target”, incorporando la predicción del tiempo de la retención, son de gran utilidad para la detección e identificación de contaminantes emergentes en agua. El uso de espectrometría de masas de baja y alta resolución, como triple cuadrupolo (QqQ), cuadrupolo tiempo de vuelo (QTOF) y trampa de iones lineal (LTQ) Orbitrap, permite el análisis exhaustivo de los contaminantes emergentes en aguas ambientales, a través del uso del análisis “target” (QqQ), “suspect” (QTOF y LTQ-Orbitrap) y “non-target” (LTQ-Orbitrap). Además, el uso complementario de la predicción del tiempo de retención hizo la detección de compuestos (especialmente para los que no estaban disponibles los patrones) mucho menos tediosa y mejoró la confianza en su identificación tentativa.

A partir del trabajo realizado en esta tesis, se pueden enumerar varias conclusiones específicas:

1. La espectrometría de masas de alta resolución, especialmente QTOF y LTQ-Orbitrap, es una técnica muy poderosa para la detección cualitativa en análisis “suspect” y “non-target” de contaminantes emergente en muestras acuáticas.
2. QTOF en el modo de MS^E facilita la detección y la identificación tentativa de los productos farmacéuticos, drogas ilícitas (PIDs) y sus metabolitos, con la información de los espectros de alta energía (que corresponden a los iones fragmento) y de baja energía (que corresponde principalmente a la molécula intacta) obtenidos en una sola inyección.
3. Se ha demostrado que hay una diferencia geográfica en la detección de PIDs, tras analizar muestras de Colombia, Italia y España. Mientras que las muestras procedentes de Italia y España eran bastante similares - irbesartán (antihipertensivo), valsartán (antagonista de angiotensina II) y dos antibióticos ofloxacina y sulfametoxazol fueron los compuestos más comúnmente identificadas - en las muestras colombianas se encontró el betabloqueante metoprolol, el regulador de lípidos gemfibrozilo, y el agente de rayos X iopromida, así como metabolitos de losartán y carbamazepina.
4. El uso del análisis “non-target” (hidden target screening) proporciona mucha información. Por otra parte, sin los patrones de referencia, permitió que numerosos compuestos pudieran ser tentativamente identificados.

5. Se ha demostrado que la predicción del tiempo de retención es una gran técnica complementaria en la detección e identificación de los PIDs y sus metabolitos.
6. Un método de predicción del tiempo de retención basado únicamente en $\log K_{ow}$ es no sólo barato y sencillo, sino que también condujo a que aproximadamente el 70-80% de todos los compuestos presentaban un tiempo de retención teórico con una tolerancia menor de 2 minutos respecto a su valor experimental.
7. Un método de predicción del tiempo de retención más complejo, basado en redes neuronales artificiales, mostró una mayor capacidad predictiva, con hasta el 95% de todos los compuestos con un tiempo de retención predicho dentro de 2 minutos respecto a su tiempo experimental. Esto permitió una mayor confianza en la identificación tentativa de los compuestos para los que los patrones no estaban disponibles.
8. La cromatografía líquida acoplada a espectrometría de masas en tándem (LC-MS/MS) con analizador de masas de triple cuadrupolo es una técnica analítica poderosa para la cuantificación de las nuevas sustancias psicotrópicas (NPS) en diferentes tipos de agua. La robustez, la sensibilidad y la selectividad de esta técnica la hacen muy adecuada para la determinación de estos compuestos a las bajas concentraciones en que se encuentran comúnmente en las muestras de matriz compleja tales como las aguas residuales urbanas.
9. Estudios adicionales de filtración son necesarios para tener en cuenta las posibles pérdidas durante el tratamiento de la muestra. Se encontró que la filtración a vacío (utilizando filtros de membrana) proporcionaba bajas recuperaciones (5%) para los tres compuestos NBOME así como la nafirona. Sin embargo, se observó que los filtros de jeringa (de PTFE, polipropileno y nylon) proporcionaban mejores recuperaciones para todos los compuestos, siendo el de Nylon de $0.45\mu\text{m}$ el mejor, ya que resultaba en recuperaciones de 98 a 105% para todos los compuestos.
10. Se estudiaron los aditivos añadidos a la fase móvil, el volumen de inyección de muestra así como los flujos de gas de desolvatación y de cono, ya que todas estas variables contribuyen a la optimización tanto de la separación y forma del pico cromatográfico como al rendimiento de ionización en espectrometría de masas.
11. La estabilidad es un parámetro importante a tener en cuenta para los estudios de epidemiología de aguas residuales en general, y para NPS en particular. La acidificación (a pH 2) fue necesaria para reducir la transformación y la degradación de todos los NPS investigados a 20°C , 4°C y -20°C durante un máximo de siete días. Si no se acidifica, las muestras deben ser analizadas dentro de las 48 horas (si se almacenan a -20°C o 4°C),

mientras que a 20°C, la N-ethylcathinone y la mefedrona mostraron una baja estabilidad degradándose dentro de las primeras 24 horas.

12. La adquisición de dos o tres transiciones SRM permite la cuantificación y la confirmación simultánea de diversas NPS.

6.3 Suggestions for Future Research

Throughout this thesis, advanced analytical techniques including high resolution mass spectrometry, tandem mass spectrometry and the complementary use of retention time prediction have shown their potential in the detection and identification of emerging contaminants in environmental waters. Frustratingly, with every result come more questions and ideas for future research, which, due either being out of the scope of this thesis or lack of time could not be undertaken. Nevertheless, several suggestions for future research are outlined below:

Screening approaches:

More information is needed on metabolites, transformation products and NPS. While the in-house databases are always being added to, this needs to continue in the future, perhaps with more emphasis on metabolites and transformation products.

To aid non-target approaches, collaboration is necessary. This could be in terms of uploading mass spectra to the online mass spectral database, MassBank, or through the sharing of in-house databases. These ideas have been floating around the NORMAN network, who have introduced a non-target specific working group for this reason, leaving the future looking bright.

Retention time prediction:

During the research stay at University of Antwerp, a retention time prediction method based on ANNs and incorporating three different columns (biphenyl, C8 and HILIC) was undertaken for 100 NPS and illicit drug reference standards. After optimizing the ANN, all three columns had a mean error within one minute: 0.51, 0.68 and 0.81 respectively, however the HILIC results are skewed due to overfitting the training set. Further work is needed on the HILIC column, while future work could include the addition of more standards to the C8 and biphenyl methods as well as its use in suspect detection and identification. When future optimization is complete, it would be very interesting to test the retention time predictor for NPS, either in wastewater or, perhaps more curiously, in toxicological samples, where “new” NPS are first identified.

A second approach could be the use of different software for retention time prediction, such as random forests or support vector machines.

Furthermore, an investigation into different molecular descriptors could be carried out. More than one thousand descriptors are available on VCCLABS' Parameter Client module. During the research stay at University of Antwerp, a brief study into the impact of different descriptors was made, but far more time is needed for a suitable and reliable study.

New Psychoactive Substances

Further to the work done on the quantification of NPS, an investigation into their metabolism could be made. The parent compound is not necessarily the most abundant and appropriate biomarker, leading to a metabolite biomarker being more useful, and perhaps leading to more compounds being confirmed in the real samples.

Further research into a direct injection approach could be made. As mentioned in Chapter 4, huge matrix effects were seen for the later eluting compounds in influent wastewater using direct injection. Alternative gradients could be investigated, perhaps performing a stepped gradient, so as to reduce the “flushing” of more non-polar contaminants at the end of the chromatographic run.

Finally, an investigation into the environmental harm of NPS and their metabolites could be made. There is currently limited information related to the environmental harm of pharmaceuticals, with even less relating to NPS.

5.4 Sugerencias para Futuras Investigaciones

A lo largo de esta tesis, se ha demostrado el potencial de las técnicas analíticas avanzadas, incluyendo la espectrometría de masas de alta resolución, espectrometría de masas tándem y el uso complementario de la predicción del tiempo de retención para la detección e identificación de contaminantes emergentes en aguas ambientales. Sin embargo, con cada nuevo resultado aparecen más preguntas e ideas para futuras investigaciones, las cuales, debido bien a estar fuera del alcance de esta tesis o por falta de tiempo no pudieron llevarse a cabo. Sin embargo, a continuación se resumen varias sugerencias para futuras investigaciones:

Aproximaciones de Screening:

Se necesita más información sobre metabolitos, productos de transformación y NPS. A pesar de que las bases de datos internas siempre están en continua actualización, en el futuro, tal vez se deba hacer más énfasis en los metabolitos y productos de transformación.

Para avanzar en aproximaciones “non-target”, la colaboración con otros grupos es vital. Esta se podría llevar cabo mediante la compartición de espectros de masas en base de datos de espectros en línea, como MassBank, o incluso mediante el intercambio de las bases de datos internas de cada grupo. Estas ideas también han sido abordadas de manera recurrente en la red NORMAN, que finalmente han cristalizado en un grupo de trabajo específico, augurando un brillante futuro en este campo.

Predicción del Tiempo de Retención:

Durante la estancia de investigación en la Universidad de Amberes, un método de predicción del tiempo de retención basado en ANNs e incorporación hasta tres columnas diferentes (bifenilo, C8 y HILIC) se llevó a cabo empleando 100 NPS y patrones de referencia de drogas ilícitas. Después de la optimización de la ANN, las tres columnas tenían una media de error por debajo del minuto: 0,51, 0,68 y 0,81, respectivamente; sin embargo, los resultados HILIC están sesgados debido a un “overfitting” en el conjunto de entrenamiento, por lo que es necesario seguir trabajando en la misma. Por otra parte, un trabajo futuro podría incluir la adición de más patrones para los métodos de C8 y bifenilo, así como su uso en la detección e identificación de sospechosos. Cuando se complete en un futuro la optimización de la ANN, sería muy interesante evaluar la predicción de tiempo de

retención para NPS, ya sea en las aguas residuales o, quizás más interesante, en las muestras toxicológicas, donde generalmente se identifican por primera vez las "nuevas" NPS.

Una segunda aproximación podría ser el uso de diferentes software para la predicción de tiempo de retención, como "random forests" o "support vector machines".

Además, se podría llevar a cabo una investigación sobre diferentes descriptores moleculares. Más de un millar de descriptores están disponibles en los módulos de Parameter Client (VCC LABS). Durante la estancia de investigación en la University of Antwerp, se hizo un breve estudio sobre el impacto de los diferentes descriptores, pero se necesita mucho más tiempo para un estudio adecuado y fiable.

Nuevas Sustancias Psicoactivas:

En relación con el trabajo realizado en la cuantificación de NPS, se podría hacer una investigación sobre su metabolismo. El compuesto original no es necesariamente el biomarcador más abundante y adecuado, lo que puede llevar a que un metabolito sea el biomarcador más útil. Este hecho permitiría tal vez que más compuestos se confirmen en las muestras reales.

También se podrían realizar investigaciones adicionales para implementar un método por inyección directa. Como se ha mencionado en el capítulo 4, se observaron grandes efectos matriz para los compuestos que eluían al final del gradiente para influente de aguas residuales al realizar la inyección directa. Se podrían investigar gradientes alternativos, p. ej. escalonados, con el fin de reducir el "lavado" de los contaminantes más apolares al final del cromatograma.

Por último, se podría hacer una investigación sobre el impacto ambiental de las NPS y sus metabolitos. En la actualidad existe información limitada relacionada con el perjuicio ambiental de los productos farmacéuticos, mucho menor para el caso de los NPS.

Curriculum Vitae

Personal Information:

Name: Richard Thomas Dacre Northcote Bade

Date of Birth: 14 September 1987, Auckland, New Zealand

Background:

Previous Education:

May 2013-: University Jaume I, Castellon, Spain. Marie Curie funded doctoral program, project entitled 'A new paradigm in drug use and human health risk assessment: Sewage profiling at the community level'.

April-May 2014: Research stay at Mario Negri Institute for Pharmacological Research, Milan, Italy

August-September 2014: Research stay at KWR Watercycle Research Institute, Nieuwegein, The Netherlands

February 2015: Research stay at KWR Watercycle Research Institute, Nieuwegein, The Netherlands

June 2015: Research stay at University of Saarland, Homburg, Germany

July-August 2015: Research stay at University of Antwerp, Antwerp, Belgium

September 2011- June 2012: University of Helsinki, Finland. Master of Science in Advanced Spectroscopy in Chemistry (ASC)

August-September 2011: University of Lille, France. ASC Summer School Course in New Methods in Solid State NMR and Related Techniques for the Analysis of Inorganic Materials.

September 2010- July 2011: University of Leipzig, Germany. Master of Science in Advanced Spectroscopy in Chemistry (ASC)

January-February 2009: University of Duisburg-Essen, Germany. DAAD Winter Course

November-December 2008: Intensive German Study Abroad Course at Die Neue Schule, Wilmersdorf, Berlin.

2006-2009: University of Auckland, New Zealand. Bachelor of Arts/ Bachelor of Science in German and Chemistry

Scholarships

2008-09: University of Auckland Summer Scholarship in Chemistry. Research on Iridabezofurans.

2008: Senior Prize in German

2009: DAAD Winterkurs Scholarship (University of Duisburg-Essen)

2009: University of Auckland Study Abroad Scholarship Languages & Literatures

2009-10: University of Auckland Summer Scholarship in Chemistry. Research on Drug-like Chemical Space.

2010: DAAD Study Scholarship for Master Course in Berlin (not taken up)

2010-2012: Erasmus Mundus Scholarship for Master of Science in Advanced Spectroscopy in Chemistry

2013: DAAD Research Grant for Doctoral Study at Technische Universität, Berlin (not taken up)

2013-: Doctoral Grant from SEWPROF MC ITN, project: 'A new paradigm in drug use and human health risk assessment: Sewage profiling at the community level' from the People Programme (Marie Curie Actions) of the European Union's 7th Framework Programme FP7/2007–2013/under REA grant agreement no. [317205].

February 2015: Short term scientific mission (STSM) funded by COST Action ES1307 "SCORE - Sewage biomarker analysis for community health assessment" for research stay at KWR Watercycle Research Institute, Nieuwegein, Netherlands.

Completion of Degrees:

Graduated with Bachelor of Arts and Bachelor of Science at the University of Auckland in May 2010.

Graduated with Master of Science in Advanced Spectroscopy in Chemistry, collaboration between University of Leipzig and University of Helsinki, in May 2012. Awarded grade Eximia cum laude approbatur for thesis (equivalent to upper 1st Class Honours), grade 5/5 (top grade) for coursework.

Additional Language Study:

Introductory Spanish as part of Master of Science in Advanced Spectroscopy in Chemistry at the University of Leipzig. October 2010 – February 2011

Mandarin 1A. Unitec Institute of Technology, Auckland, New Zealand. August – November 2012.

Spanish 1A. School of Languages, Castellon. September 2013-May 2014

Work Experience, Practical Training:

November 2008-February 2009: University of Auckland Summer Scholarship in Chemistry. Research on Iridabenzofurans. Supervisor: Associate Professor L. James Wright.

November 2009-February 2010: University of Auckland Summer Scholarship in Chemistry. Research on Drug-like Chemical Space. Supervisor: Dr Jóhannes Reynisson

April 2010-September 2010: Internship in New Product Development at Douglas Pharmaceuticals in Auckland, New Zealand. Supervisor: Henele Schaaf

November 2011-April 2012: Master's Thesis: Characterisation of novel multicomponent reaction products using 2D NMR. Supervisor: Professor Kristiina Wähälä.

Presentations:

Oral:

6 August 2009: Goethe Society Lecture, University of Auckland. "Impressions of the DAAD Winterkurs at University of Duisburg-Essen"

16 May 2012: University of Helsinki, Finland. Organic Chemistry Day. Presentation on "Characterisation of Novel Multicomponent Reaction Products using Two Dimensional NMR." Accepted award for best international presentation.

12 June 2012: Technische Universität Berlin, Germany. Invited lecture on "Characterisation of Novel Multicomponent Reaction Products using Two Dimensional NMR".

1 September 2015: NORMAN Cross-Working Group Activity on Non-target Screening, Rhodes, Greece. Presentation "Progress with retention time prediction/ index". Presented with N. Thomaidis (University of Athens, Greece) and M. Krauss (UFZ, Leipzig, Germany).

11-15 October 2015: Testing the Waters 2015, Ascona, Switzerland. "The Development and Use of Retention Time Prediction in the Identification of Emerging Contaminants in Environmental Waters" R. Bade, L. Bijlsma, L. P. Barron, T. H. Miller, J. V. Sancho and F. Hernandez

28-30 October 2015: XV Scientific Meeting of the Spanish Society of Chromatography and Related Techniques (SECyTA2015), Castellon, Spain. "Prediction of Chromatographic Retention Time using Artificial Neural Networks for the Identification of Emerging Contaminants in Environmental Waters" R. Bade, L. Bijlsma, L. P. Barron, T. H. Miller, J. V. Sancho and F. Hernandez

22 December 2015: Invited Presentation at Escuela de Empresarios (EDEM), Valencia, Spain: "SEWPROF: a collaborative approach to sewage-based epidemiology"

6 January 2016: The 7th Annual Multidimensional Chromatography Workshop, Toronto, Canada. "The Use of Artificial Neural Networks for the Prediction of Chromatographic Retention Time in the Identification of Pharmaceuticals and Illicit Drugs Metabolites in Environmental Waters" R. Bade, L. Bijlsma, L. P. Barron, T. H. Miller, J. V. Sancho, M. Ibáñez, N. Thomaidis and F. Hernandez

26 May 2016: Invited presentation: Taller Redonda para Doctorandos (Workshop for Doctoral students), University Jaume I, Castellon, Spain. "Experiences of participating in a Marie Sklodowska Curie Action"

Poster:

25-28 June, 2013: ICCE 2013: 14th EuCheMS International Conference on Chemistry and the Environment (ICCE), Barcelona, Spain. "Modern analytical methodology based on UHPLC-MS/MS for the sensitive determination of illicit drugs in urban wastewaters with emphasis on amphetamine and derivatives" Authors: L. Bijlsma, R. Bade, E. Beltran, J.V. Sancho, F. Hernández.

3-5 September, 2015: CEST 2015: 4th International Conference on Environmental Science and Technology, Rhodes, Greece. "The Use of Artificial Neural Networks for Retention Time Prediction in the Identification of Metabolites and Transformation Products in Environmental Waters" Authors: R. Bade, L. Bijlsma, L. P. Barron, T. H. Miller, J. V. Sancho and F. Hernandez

11-15 October 2015: Testing the Waters 2015, Ascona, Switzerland. "The use of advanced deconvolution software in the post-target and non-target screening of wastewater samples by High Resolution Mass Spectrometry" R. Bade, A. Causanilles, L. Bijlsma, E. Emke, P. de Voogt, J. V. Sancho, F. Hernández

27 October 2015: VII National Meeting of the Spanish Society of Mass Spectrometry (SEEM2015) "Advanced Deconvolution Software in the Suspect and Non-Target Screening of Wastewater Samples" R. Bade, A. Causanilles, L. Bijlsma, E. Emke, P. de Voogt, J. V. Sancho, F. Hernández

Conferences and Workshops:

December 2011: Third International Workshop on Analysis of Chemical Warfare Agents, Helsinki, Finland.

6-8 May, 2013: Testing the waters: first international multidisciplinary conference on detecting illicit drugs in wastewater, Lisbon, Portugal.

25-28 June, 2013: ICCE 2013: 14th EuCheMS International Conference on Chemistry and the Environment (ICCE), Barcelona, Spain.

25 June, 2013: ICCE 2013 Satellite Event: Illicit Drugs in Wastewaters, Barcelona, Spain.

10 July, 2013: I Workshop ISIC 2012/016- Envi Food, Castellon, Spain.

11 July, 2013: Seminar “Emerging environmental contaminants: occurrence and implications” by Dr. Pim de Voogt and Dr. John Parsons, Castellon, Spain.

9-10 September, 2013: SEWPROF Training Course 2: Sampling, sample handling, storage and sample preparation. Oslo, Norway.

10-11 April, 2014: SEWPROF Training Course 3: Analytical techniques for biomarkers analysis. Utrecht, Netherlands.

17-18 September, 2014: SEWPROF Training Course 4: Assessing human health and lifestyle by sewage epidemiology. Milan, Italy.

4 November, 2014: “Introduction to suspect and non-target screening with HR-MS(MS)” by Dr. Emma Schymanski, Castellon, Spain.

5 November, 2014: “Improving Identification of Unknowns with MassBank” by Dr. Emma Schymanski, Castellon, Spain

29-30 April, 2015: SEWPROF Training Course 5: Method Validation and Quality Control. Castellon, Spain.

Publications/Posters:

2010: Richard Bade, Ho-Fung Chan and Jóhannes Reynisson, Characteristics of known drug space. Natural products, their derivatives and synthetic drugs. *European Journal of Medicinal Chemistry* 2010 45 (12) 5646-5652.

2012: Richard Bade (ed.) *German Perspectives on New Zealand and the Pacific from the Eighteenth Century to the Present Day: Four Research Essays*. University of Auckland Research Centre

- for Germanic Connections with New Zealand and the Pacific. Auckland, 2012. (= Research Centre Working Papers Series Volume 7) ISBN: 0-9582345-9-0
- 25-28 June, 2013: Poster: “Modern analytical methodology based on UHPLC-MS/MS for the sensitive determination of illicit drugs in urban wastewaters with emphasis on amphetamine and derivatives” Authors: L. Bijlsma, R. Bade, E. Beltran, J.V. Sancho, F. Hernández. ICCE 2013: 14th EuCheMS International Conference on Chemistry and the Environment (ICCE), Barcelona, Spain.
- 10-14 August, 2014: Poster: “Screening of illicit and licit drugs in waters from Colombia making use of liquid chromatography-hybrid Quadrupole-time of flight mass spectrometry” Ana María Botero-Coy, L. Bijlsma, R. Bade, M. Ibáñez, M. C. Bustos, J. Vargas-Martínez, R. J. Rincón, A. Moncayo, J. V. Sancho, F. Hernández. 248th American Chemical Society National Meeting, San Francisco, USA.
- 16 October, 2014: Oral Presentation: “Advanced analytical approaches to investigate the presence of undesirables from feed to fish. Atlantic salmon and gilthead sea bream as a case study.” J. Náchter-Mestre, R. Serrano, E. Beltrán, C. Boix, L. Bijlsma, R. Bade, T. Portolés, M. Ibáñez, V. Karalazos, J. Silva, M. H. G. Berntssen, J. Pérez-Sánchez, F. Hernández. Aquaculture Europe 2014: International Conference and Exposition, San Sebastian, Spain.
- 2014: Investigation of illicit drugs and pharmaceuticals in waters by liquid chromatography-high resolution mass spectrometry F. Hernández, M. Ibáñez, R. Bade, L. Bijlsma, J.V. Sancho, Trends in Analytical Chemistry 2014, 63, 140-157.
- 2015: Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water R. Bade, L. Bijlsma, J. V. Sancho, F. Hernández, Talanta 2015, 139, 143-149.
- Screening of pharmaceuticals and illicit drugs in wastewater and surface waters of Spain and Italy by high resolution mass spectrometry using UHPLC-QTOF MS and LC-LTQ-Orbitrap MS R. Bade, N. I. Rousis, L. Bijlsma, E. Gracia-Lor, S. Castiglioni, J. V. Sancho, F. Hernández, Analytical and Bioanalytical Chemistry 2015, 407, 8979-8988.

LC-QTOF MS screening of more than 1000 licit and illicit drugs and their metabolites in wastewater and surface waters from the area of Bogotá, Colombia F. Hernández, M. Ibáñez, A. M. Botero-Coy, R. Bade, M. Bustos-Lopez, J. Rincon, A. Moncayo, L. Bijlsma, *Analytical and Bioanalytical Chemistry*, 2015, 407, 6405-6416.

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 R. Bade, L. Bijlsma, T. H. Miller, L. P. Barron, J. V. Sancho, F. Hernández, *Science of the Total Environment*, 2015, 538, 934-941.

2016: Critical review on the stability of illicit drugs in sewers and wastewater samples A-K McCall, R. Bade, J. Kinyua, F. Y. Lai, P. K. Thai, A. Covaci, L. Bijlsma, A.L.N. van Nuijs, C. Ort, *Water Research*, 2016, 88, 933-947.

UHPLC-QTOF MS Screening of Pharmaceuticals and their Metabolites in Treated Wastewater Samples from Athens M. Ibáñez, V. Borova, C. Boix, R. Aalizadeh, R. Bade, N.S. Thomaidis, F. Hernández *Journal of Hazardous Materials*, 2016 ACCEPTED

3-Fluorophenmetrazine, a fluorinated analogue of phenmetrazine: studies on in vivo metabolism in rat and human, in vitro metabolism in human CYP isoenzymes and microbial biotransformation in *Pseudomonas Putida* and wastewater using GC and LC coupled to (HR)-MS techniques M. Mardal, B. Miserez, R. Bade, T. Portolés, M. Bischoff, F. Hernandez, M. R. Meyer, *Journal of Pharmaceutical and Biomedical Analysis*, 2016, ACCEPTED

Facilitating high resolution mass spectrometry data processing for screening of environmental water samples: An evaluation of two deconvolution tools R. Bade, A. Causanilles, E. Emke, L. Bijlsma, J. V. Sancho, F. Hernández, P. de Voogt *Science of the Total Environment*, 2016 UNDER REVISION

Ultra-high performance liquid chromatography-tandem mass spectrometry determination of synthetic cathinones and phenethylamines in wastewater of eight European cities R. Bade, L. Bijlsma, J. V. Sancho, J. A. Baz-Lomba, S. Castiglioni, E. Castrignanò, A. Causanilles, E. Gracia-Lor, B. Kasprzyk-Hordern, J. Kinyua, A.-K. McCall, A.L.N. van Nuijs, C. Ort, B. G. Plósz, P. Ramin, N. I. Rousis, Y. Ryu, K. Thomas, P. de Voogt, F. Hernández, 2016, SUBMITTED FOR PUBLICATION

Acknowledgements

Carrying out doctoral studies in a foreign country on the other side of the world has been a daunting task, but it has been made a lot easier by my group here in Castellon as well as my friends and family scattered around the world.

First of all, I would like to thank everyone involved with enabling me to come here to study, from the first interview with Felix, Robert and Elena, to the various “issues” involved with getting me over here, and the great welcome I received from the IUPA. While my Spanish learning was continually interrupted with my secondments, I did learn some: quisiera expresar mi agradecimiento a todo el IUPA, por su ayuda y consejos. Cuando tenía una duda, todos estaban dispuestos y disponibles para ayudar. Mención especial querría tener con Vima: sin ella no hubiera practicado mi castellano... y ella su inglés. He aprendido mucho con todos y deseo a cada una todo lo mejor para el futuro.

Relating to my research, I would like to thank my supervisors at IUPA, Felix and Juanvi. They gave me flexibility to work independently, however, whenever I had any concerns, they were always there for me, no matter where in the world we were! They were a fountain of knowledge and encouraged me throughout all my work, and I give my heartfelt thanks to them for the opportunity to work with such brilliant minds. I would also like to thank my external SEWPROF supervisor, Pim. Although he came into the supervisory role quite late, working with such an amazing person made my research and output all the more better.

I would also like to thank Robert. He has really morphed from a colleague to a mentor and friend. He was the first person I met upon arriving in Spain, and has (in unofficial capacity as my pseudo-supervisor) helped me immensely in all my work, always being my go-to man. Words cannot give justice to my thanks to him.

I would like to thank everyone within SEWPROF. All the ESRs and ERs: Pepo, Ann-Kathrin, Yeonsuk, Stefania, Ana, Juliet, Marie, Nikos, Erika, Pedram, Janelle, Emma, Bram and Zhugen; and supervisors: Kevin, Christoph, Jørgen, Pim, Adrian, Markus, Sara, Barbara, Benedek, Guido and John. It has been a real pleasure getting to know all of you throughout the training courses and conferences we have attended as well as the collaborative research we have undertaken on various secondments and I hope that in the future we can continue to work together.

One of the biggest advantages of being a part of a Marie Curie doctoral program (and one of the reasons I applied for this project) is the mobility. Without a doubt, I am privileged to have been able to travel to other parts of Europe to work with some fantastic groups – at the Mario Negri Institute, KWR, University of Saarland and the University of Antwerp. Special mention has to go to the supervisors I had during my secondments at Mario Negri Institute (Sara and Ettore), KWR (Pim and Erik), University of Saarland (Hans) and the University of Antwerp (Adrian and Alex).

Living so far away from what I knew was initially very difficult, and I could not have made it without Xing ah 我很爱你 I think many knew of my troubles adjusting to a new country, climate, language and mentality, but, honestly after three years here, I do think I will really miss it.

Finally, I'd like to thank my family. Even though they were so far away and may not have understood quite what I was doing, they never stopped encouraging me.

At some stages of my doctoral work, when the smell of wastewater and/or the screen of my computer was making me sick... I was thankful for the weekend. So, special mention has to go to the people who helped me through some of those tough times: the Warriors, Black Caps, All Blacks, Liverpool and (after arriving in Spain) Villarreal, Barcelona and the local C. F. San Pedro were there to help keep my mind off work.