UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÁ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI

INDIRIZZO SCIENZE CHIMICHE

CICLO XXVIII

ANALYTICAL PROTOCOLS BASED ON HIGH-RESOLUTION MASS SPECTROMETRY FOR CHARACTERIZING EMERGING CONTAMINANTS AND THEIR DEGRADATION PRODUCTS IN FOODSTUFF AND ENVIRONMENT.

Direttore della Scuola: Ch.mo Prof. Antonino Polimeno

Supervisore: Ch.mo Prof. Sara Bogialli

Dottorando: Bortolini Claudio

Anno 2013-2016

Ama e fa ciò che vuoi.

Table of contents

1 Abstract

In this PhD thesis, the capability of analytical systems based on high-resolution mass spectrometry (HRMS) has been investigated for the determination of emerging contaminants in environmental matrices and foodstuff. Since the molecular structures of the emerging contaminants could be know as well as unknown, target, suspect and non-target analyses have to be developed in order to propose a "mass-based" advanced screening. Attention has been focused on the scale-up process in the identification confidence by developing different specific protocols.

Two protocols based on HPLC/Q-TOF-MS have been developed for the simultaneous screening and confirmatory analysis of target and non-target cyanotoxins in freshwater intended for human consumption, PDE-5 inhibitors and analogues in food supplements marked as erectile dysfunction remedies. Both protocols have been optimized with the aim to obtain HRMS data of pseudomolecular ions and fragmentation patterns in tandem MS mode. In-house databases were implemented to simplify the data treatment.

The application of these protocols in "non-target screening" mode has been attempted in real samples and in the frame of a collaborative trial organized by European NORMAN foundation as regard as the analysis of water contaminants. The exercise was complex and time consuming, and it has highlighted the strengths and weaknesses of the developed protocols.

The crucial step in non-target screening was the assignment of reliable molecular formula to the *m/z* values. A specific workflow based on direct infusion and HRMS analysis by using an Orbitrap™ mass spectrometer has been developed for the characterization of $PM_{2.5}$ organic fraction. The automatization of the data treatment using Mathematica based algorithms was accomplished for studying the chemical composition of $PM_{2.5}$ organic fraction. Contextually, the possible use of the Atmospheric Pressure Photoionization source for characterizing $PM_{2.5}$ organic fraction has been investigated on real samples.

1.1 Riassunto

In questa tesi di dottorato, le possibilità dell'uso della spettrometria di massa ad alta risoluzione (HRMS) sono state indagate nella determinazione di contaminanti emergenti in matrici ambientali ed alimentari. Dal momento che le strutture molecolari dei contaminanti emergenti potrebbero essere ancora sconosciute, la loro determinazione richiede analisi di tipo *target*, di composti sospetti e *non-target* devono essere sviluppati al fine di proporre una metodologia di screening avanzata basata sulla spettrometria di massa. L'attenzione è stata focalizzata sul processo di *scale-up* nella confidenza di identificazione, sviluppando protocolli analitici specifici.

Due protocolli per la simultanea analisi *target* e di composti sospetti, basati sulla piattaforma HPLC-Q-TOF, sono stati sviluppati ad applicati nell'analisi di cianotossine in acqua dolce destinata al consumo umano e inibitori del PDE-5 negli integratori alimentari venduti come rimedi per la disfunzione erettile. Entrambi i protocolli sono stati ottimizzati con lo scopo di ottenere dai dati mass-spettrometrici in modalità tandemMS, gli ioni pseudomoleculari e lo spettro di frammentazione. Librerie sono state sviluppate e implementate per semplificare il trattamento dei dati.

La possibile applicazione di questi protocolli nell'analisi di tipo *non-target* è stata tentata su campioni reali nell'ambito di una prova collaborativa organizzata dall'associazione Europea NORMAN, riguardante l'analisi di contaminanti nell'acqua. L'esercizio è stato complesso e richiedente molto tempo, e ha messo in evidenza i punti di forza e di debolezza dei protocolli sviluppati.

Il passaggio cruciale nell'analisi *non-target* è l'assegnazione di formule molecolari veritiere ai valori *m/z*. Un *workflow* di analisi basato sulla infusione diretta del campione e l'acquisizione di massa utilizzando un Orbitrap ™ è stato sviluppato e automatizzato utilizzando algoritmi basati sul linguaggio di programmazione Mathematica, per studiare la composizione chimica della frazione organica del $PM_{2.5}$. Contestualmente, l'eventuale uso della fotoionizzazione a pressione atmosferica (APPI) per la caratterizzazione frazione organica del $PM_{2.5}$ è stata indagata su campioni reali.

2 List of Abbreviations

ACN	Acetonitrile
Adda	3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid
ANA-a	Anatoxin-a
ANP	Anabaenopeptin
APCI	Atmospheric pressure chemical ionization
APPI	Atmospheric pressure photoionization
B(α)P	Benzo[α]pyrene
BMAA	β-*N*-methylamino-L-alanine
BPC	Best peak chromatogram
CE	Collision energy
CID	Collision induced dissociation
CRM	Certified reference material
CYL	Cylindrospermopsin
DBE	Double bond equivalent
EC	Emerging contaminant
EIC	Extracted ion current
ELISA	Enzyme-linked immune assay
ESI	Electrospray
FA	Formic acid
GC	Gas chromatography
HPLC	High-pressure liquid chromatography
HRMS	High-resolution mass spectrometry
HWHM	Full width at half maximum
KM	Kendrick mass
KMD	Kendrick mass defect
LC	Liquid chromatography
LOD	Limit of detection and quantification
LOQ	Limit of quantification
LTQ	Linear Ion Trap
m/z	Mass to charge ratio
MC	Microcystin
MME	Mass measure error

MS	Mass spectrometry
MS/MS	Tandem Mass Spectrometry
MW	Molecular weight
NL	Noise level
NMR	Nuclear magnetic resonance
NOD	Nodularin
OSc	Oxidation state of the carbon
PAH	Polycyclic aromatic compounds
PDE-5	Phosphodiesterase type 5
PFPA	Pentafluoropropionic acid
PM	Particulate matter
POA	Primary organic aerosol
PTFE	Polytetrafluoroethylene
Q-TOF	Quadrupole-time-of-flight
RSD	Relative standard deviation
RTI	Retention time index
S/N	Signal to noise ratio
SD	Standard deviation
SOA	Secondary organic aerosol
STX	Saxitoxins
TFA	Trifluoroacetic acid
TIC	Total ion current
VK	Van Krevelen
VOC	Volatile organic compound

3 Introduction

The main purpose of an analytical chemist is to provide information about chemical constituents in a sample. In the case of quantitative information are request, the question he/she has to pone himself or herself is: "How much analyte is in the sample?". A good answer obviously has to provide, before the quantitative value, all the unquestionable information indicating that the result refers to the considered analyte. When only qualitative evidences are requested, two different questions are anyway possible: "Is the analyte in the sample?" or "What is present in the sample?".

The analytical approach to be followed depends on the different concepts under these two questions: while in the first case, a target analysis will be performed, in the second case the commonly defined "screening analysis" can be attempted [1].

The analysis of emerging contaminants is a topic where screening analysis is pivotal.

3.1 Emerging contaminants

With the term "emerging contaminants" (ECs) or "contaminants of emerging concern" we identify substances that are recently taken into account from the scientific community because they represent a potential risk factor for human health or environment. Drewes and Shore 2001 [2]defined ECs as chemicals that have recently been shown to widely occur in water resources. Although adequate data to determine the correlated risk do not yet exist, these are anyway identified as potential toxicants. This definition is limited to water pollution, whilst a proper exhaustive categorization of ECs has to be extended to other compartments or matrices.

Starting from the previous definition, we would like to propose a definition of ECs substances that exhibit a somehow identified factor of risk, or whose hazard will be a factor of risk in the future for human health or environment.

From our definition, may be considered ECs compounds showing:

- Known exposure route + newly identified hazard.
- Known hazard + new exposure route.
- Known hazard + known exposure + increasing in human's sensitivity.

All the cited factors contribute to the risk associated to these compounds. This definition pays particular attention also on substances that might be a risk in the future. Running examples are far to be rare, e.g. human's sensitization to allergens.

Several compounds or class of compounds are currently been recognized as ECs in the water cycle: Perfluorinated compounds (PFCs) [3], brominated compounds and flame-retardants [4], personal care products (PCPs), (pharmaceuticals, hormones, cosmetics and sunscreens), drugs of abuse[2], disinfection by-products, nanomaterials[5], artificial sweeteners [6], together with their correlated transformation products [2, 7]. Other substances of concern are benzotriazoles, used as anticorrosive and for silver protection in dishwashing liquids, the naphthenic acids arising from crude oil, synthetic musk fragrances, prions, which are infectious particles composed of a protein in a misfolded form, and ionic liquids replacing traditional solvents used in industry [8].

Among the already known or potential ECs, three classes of compounds have been specifically taken into consideration in this PhD thesis: cyanotoxins, $PM_{2.5}$ and a class of pharmaceuticals, the phosphodiesterase-5 inhibitors (PDE-5 inhibitors).

3.1.1 Cyanotoxins

Cyanobacteria are worldwide spread prokaryotic organisms present on Earth since early stages of life. Their role in the evolution of the Earth is of key importance, because these photosynthetic organisms, firstly released oxygen in the atmosphere. Cyanobacteria are ubiquitous present in eutrophic water reservoirs and they are notably associated to the unpleasant odour that may occur in drinking water, due to their ability in producing compounds as 2-methylisoborneol and geosmin [9]. However, the main issue of concern in surface waters is related to the ability of several species of cyanobacteria in producing secondary metabolites, toxic for several organisms including humans [10].

These cyanotoxins could be produced, when the biomass of cyanobacteria grows drastically in a short period, thus causing dense blooms on the water, and often with a remarkably colouring of the water surface e.g. “the red tide” of the *Planktothrix rubescens*. Although the presence of more than one genotype of cyanobacterium is rare [11], a co-blooming of different species could be experienced. In Italy the incidence of potentially toxic algal blooms is very high, with more than 60 basins interested [12].

Harmful cyanobacterial blooms are regulated by both genetic and environmental factors. Among the latter ones, water temperature is the most important, as many species of cyanobacteria prefers warm water (more than 25 °C). Incidentally, global warming is considered an indirect cause of the increasing occurrence of toxic algal

blooms [13][14]. The water concentration of macronutrients is also important: cyanobacteria efficiently grow in lentic aquatic ecosystems with relatively high concentrations of primary nutrients as nitrogen, phosphorus, and carbon [15], with a known correlation with the nitrogen/phosphorus concentration ratio[16]. The increasing release in the environment of nutrients coming from farming and agricultural activities, together with nutrient accumulation promoted by the long residence times in lakes and reservoirs, feed cyanobacteria blooms [16].The last environmental factor influencing cyanobacterial bloom is the light exposure. Most species can effectively make photosynthesis only in a limited range of light quality, intensity, and duration [9].

The Cyanotoxins are classified emerging contaminants since only few congeners are well known in terms of toxicity and in parallel to the discovery of new congeners, also new exposure routes have been identified.

3.1.2 Cyanotoxins classification

The potential toxicity of cyanobacteria is related to the biosynthesis of the harmful secondary metabolites produced. About 40 of the 150 known phytobacterium genera are able to produce toxins, classified according to their mode of action primarily into hepatotoxins, neurotoxins and skin irritants [9].

3.1.2.1 Hepatotoxic algal metabolites

Microcystins (MCs) and cylindrospermopsin [17-19] are the most diffused hepatotoxins in freshwaters, produced by various species within the genera *Microcystis, Anabaena, Oscillatoria, Nodularia, Nostoc, Cylindrospermopsis*, and *Umezakia*. Their occurrence has been reported in Asia, Europe, North Africa, North America and Scandinavian countries. MCs are monocyclic heptapeptides with relatively low molecular weight. More than 100 congeners of MCs with a general structure (-D-Ala^{1}-X^{2}-D-MeAsp^{3}-Z^{4}-Adda^{5}-D-Glu^{6}-Mdha^{7}-) are known. This wide number of congeners is primarily due to the variability in composition of the amino acidic residues in positions X end Z (Figure 1). For example, the most studied MC, which has leucine (initial L), and arginine (initial R) in position 2 and 4 respectively, is identified as MC-LR.

Hepatotoxins show their toxicity in the liver where are quickly concentrated. The effect is dose dependent and mainly related to their interaction with the protein phosphatases (PP1A and PP2A), causing the inhibition of enzymatic activity with cell necrosis

followed by massive haemorrhages and death. These adverse effects seem to be ascribed to the unusual Adda amino acid, almost invariably shared by all MC variants.

Nodularins (NODs) are mainly associate with booms of *N. spumigena*, which occurrence have been reported Australia, New Zealand and the Baltic Sea [9]. NODs are cyclic pentapeptides structurally similar to MCs, including the Adda moiety but with only one variable amino acid. So far, nine variants have been identified, the most common being NOD-R with arginine as variable amino acid.

Whilst NODs are potential tumour promoters with hepatotoxic toxicity similar to MCs, no human intoxication has been reported and reliable toxicological data are not quoted.

Cylindrospermopsin (CYL) was initially described as a tropical toxin because occurred in Australia, New Zealand and Thailand. Anyway, recent reports in temperate areas, such as Italy [19], Germany and France [8] have widened its ecological habitat. CYL was initially named from the algal specie producer, the cyanobacterium *Cylindrospermopsis raciborskii*, but nowadays *Aphanizomenon ovalisporum*, *Raphidiopsis curvata* and *Umezakia natans* have been described to perform the biosynthesis of this toxin.. CYL is a highly polar tricyclic alkaloid (Figure 1) with a guanidine moiety along with a uracil, which is described as potentially responsible for its toxicity. After ingestion, the toxin mainly affects the liver via the irreversible inhibition of protein synthesis leading to cell death. To date, two other variants have been reported, *i.e.* 7-epicylindrospermopsin and the non-toxic deoxycylindrospermopsin.

Anabaenopeptins (ANPs) is another class of hepatotoxic algal metabolites. They are unique cyclic peptides that have the common cyclic peptide moiety linked with Tyr, Arg, Lys, and Phe through an ureido bond. The most representative congeners of this class are the Anabaenopeptin-A and Anabaenopeptin-B, but several congeners are reported in literature. [20].

The microginin FR1 was the first congener of the class isolated from a water bloom of a German lake [21]. Microginin FR1 structure is as linear peptide containing β-amino-α-hydroxy-decanoic acid (Ahda), alanine, *N*-methyl-leucine, and two tyrosine units (Ahda-Ala-*N*-Me-Leu-Tyr-Tyr). Microginin FR1 had angiotensin-converting enzyme inhibitory activity. Recently several microginin congeners have been identified and reported [22].

3.1.2.2 Neurotoxic algal metabolites

Anatoxins are well described neurotoxic algal metabolites. Anatoxin-a (ANA-a) and homoanatoxin-a (Figure 1) are the two main neurotoxic alkaloids produced by *Anabaena*, *Aphanizomenon* and *Planktothrix* cyanobacteria, whose occurrence was reported in USA, Africa, Asia and Europe [9, 17, 23].

ANA-a induces paralysis of the organism by interacting with acetylcholine receptors with the consequent death by respiratory arrest. The LD_{50} in mice is 375 μg/kg 24 h after the intra peritoneal injection. Animal poisoning by ANA-a causes vomit, convulsion and respiratory arrest. ANA-a(s) is the phosphate ester of a cyclic N-hydroxyguanine, with similar toxic behaviour of ANA-a, which has been identified associated to *Anabaena* strains in restricted areas of United States, Scotland, Denmark and Brazil [9].

The other main class of algal neurotoxins is saxitoxins (STXs), which have been detected in freshwaters of Australia and USA. Saxitoxins are biosynthesized by *Anabaena circinalis* and *Aphanizomenon flos-aquae,* but also *Lyngbya wollei* and *C. raciborskii* are known to be able to express these compounds. Saxitoxins are tricyclic compounds that can be non-sulphated, singly sulphated or doubly sulphated. These toxins can persist over 90 days in freshwater and can be converted into more toxic variants by high temperatures. As other neurotoxic metabolites, STXs are paralytic shellfish poisons, blocking sodium ion channels in membrane of nerve axons, and finally inducing death due to respiratory failure [17].

β-*N*-methylamino-L-alanine (BMAA) is a cyanotoxin recently identified in England, Peru South Africa, China and USA. BMAA is a non-protein amino acid acting on glutamate receptors and blocking motor neurons. In addition, BMAA could also cause intra-neuronal protein misfolding associated to neurodegeneration, and some studies connect the exposure to BMMA to the amyotrophic lateral sclerosis. This toxin has been reported to be produced by all known groups of cyanobacteria that possess genes encoding for cysteine synthase-like enzyme and methyl transferase, both involved in the BMAA biosynthesis. Anyway, several toxicological data are considered not reliable, because of possible misidentifications [20].

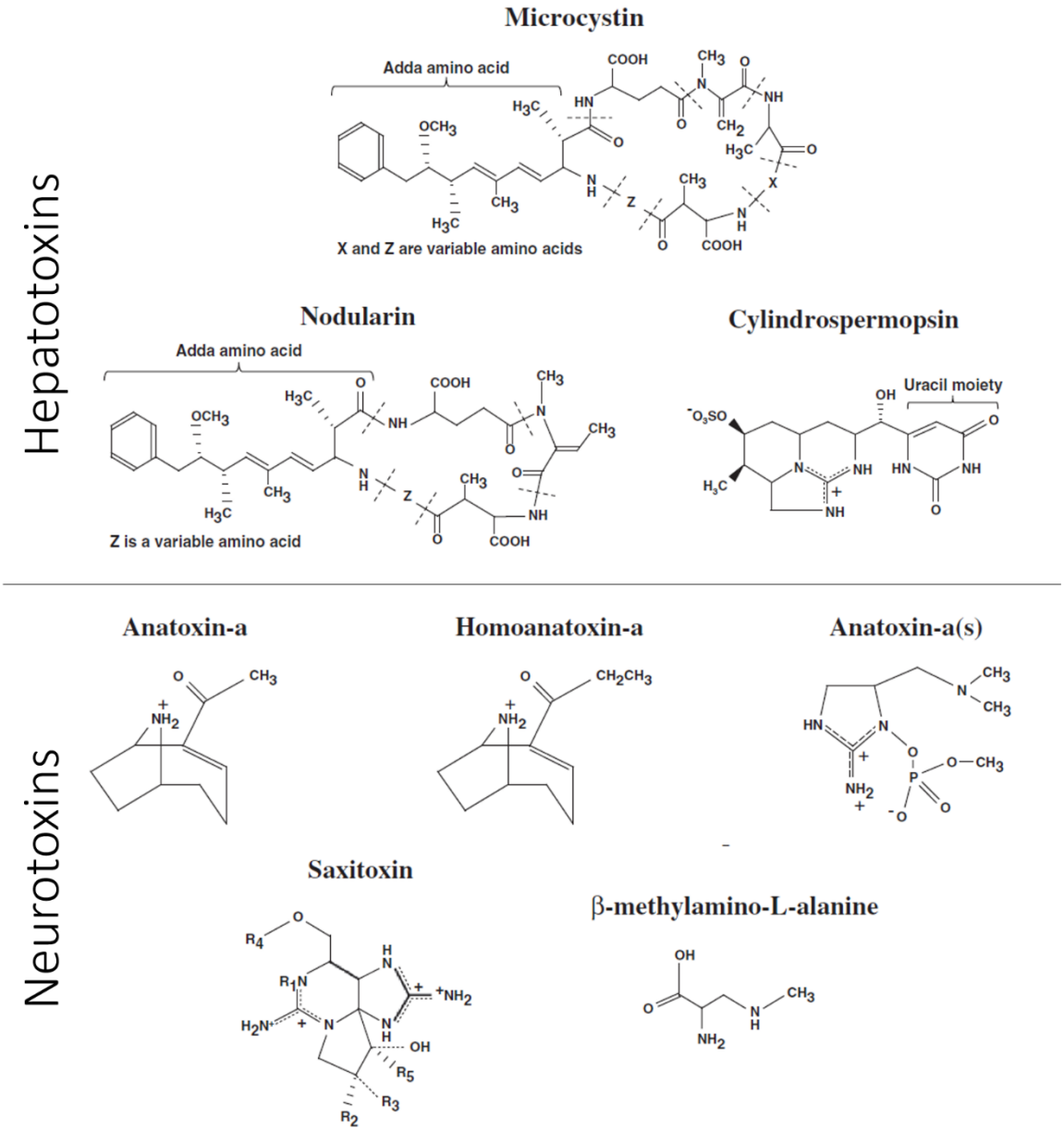

Figure 1. Widespread hepatotoxic and neurotoxic algal toxins.

3.1.2.3 Cyanotoxins exposure and regulation

Humans could be chronically exposed to cyanotoxins *via* contaminated drinking water [9, 18, 24] or food, including dietary supplements. Assumption by drinking water could affect a large portion of population of the area served from a contaminated reservoir. Another exposure route is the possible contact through dermal and accidental inhalation/ingestion during recreational activities in waters subjected to a toxic bloom. Last route of exposure is the ingestion of cyanobacteria-based food ingredients or shellfish which previously bio-accumulated toxins through filtration of contaminated water.

The World Health Organization (WHO)[25] recommended a provisional guideline value of 1 µg/L of MC-LR equivalents in drinking water, and regulatory values were currently set in several countries specifically for MC-LR or MC-LR equivalents, anatoxins, CYL (0.1 to 15 µg/L) and saxitoxins (3 µg/L). Italy together with France and Turkey will be one of the first European countries to adopt a regulatory value of 1.0 µg/L for total content (intracellular +extracellular) of MCs in drinking water, intended as sum of all congeners that can be quantified. This very conservative approach with regard to health protection was inspired by a case study [12], and by a recent evaluation of the relative protein phosphatase (PP) inhibitory ability of several MCs variants compared to the MC-LR congener. Toxicological information relative to other cyanobacterial oligopeptides is not yet reliable, and no indication is currently available from WHO about their risk assessment.

New Zealand is the only country regulating simultaneously MCs, NOD, CYL, ANA-a, homoANA-a, ANA-a(s) and saxitoxins, due to the large number of case reports described in this country. However, emerging cyanotoxins like BMAA, aplysiatoxins and lyngbyatoxins are not considered, probably because the lack of data did not allow the calculation of a guideline [9].

3.1.3 PDE-5 inhibitors in food supplements

Food supplements and herbal remedies for the treatment of erectile dysfunction and for increasing sexual performance are getting from year to year more widespread [26, 27].

Various factors are responsible for the increased demand of these products:1) certainly consumers perceive natural products much safer and healthier than drugs containing synthetic active ingredients; 2) these products are available also without prescription outside the official health system, e.g. in herbalist's shop, sex-shops and online market; 3) these products are often cheaper than official drugs [26]. As the Figure 2 shows, sexual performance enhancement remedies represent the most counterfeit products recognized on the market.

PDE-5 Inhibitors have a known toxicity but the new exposure route represented by the adulteration of food supplement classifies them as ECs.

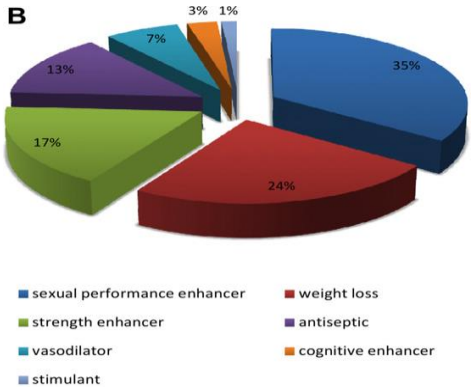

Figure 2. Notifications for dietetic foods, food supplements, and fortified foods from January 2003 to July 2010. The proportion of different contaminants by effect, listed in decreasing order. Adapted from [26].

Among the PDE-5 inhibitors, sildenafil citrate (Viagra®), tadalafil (Cialis®) and vardenafil (Levitra®) are approved in Europe and USA, udenafil (Zydena®) in South Korea and Malaysia, mirodenafil (Mvix®) in South Korea, and lodenafil carbonate (Helleva®) in Brazil [28]. These active principles are used in the official drugs and are also used in counterfeiting of food supplements alongside many possible analogues. A Dutch study carried out on 538 illicit erectile dysfunction remedies collected by the official European health agencies showed that 98% contained a PDE-5 inhibitor such as sildenafil (72%) and tadalafil (14%), while vardenafil was found in only 2% of the products. The authorized active principles are the most used in counterfeit food supplement because of the facility in the reparability and the sure effects on the sexual enhancement. Structural analogues of the official drugs have been also detected in adulterated food supplements. Most of the analogues identified are obtained from synthesis of sildenafil by modifying the reaction intermediate, so that hundreds of variants are possible. The most diffused analogues of sildenafil, homosildenafil and hydroxyhomosildenafil (Figure 3), derive from a modification of the pyperazine ring.

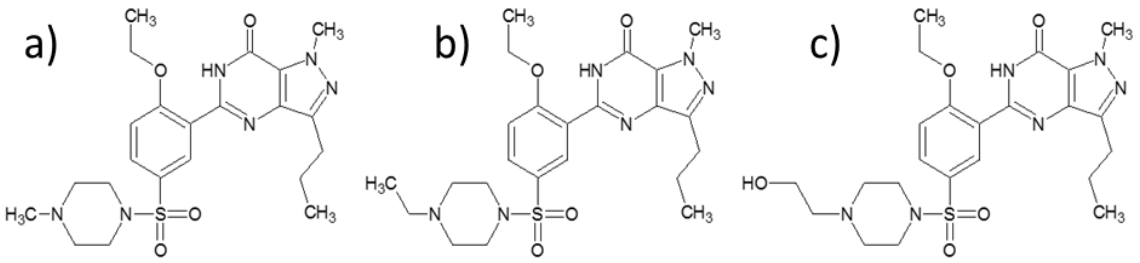

Figure 3. a) sildenafil; b) homosildenafil; c) hydroxyhomosildenafil.

Sildenafil thio-derivatives have been synthesized by heating with P_2S_5, and have the same possible variety in analogues of their oxo-counterparts, with respect to whose they are described as more powerful.

Tadalafil (Figure 4-a) is known for the shorter synthesis and the advantage to have a different pharmacokinetic with respect to sildenafil. In fact, it exhibits a much longer time window (36 hours) than sildenafil (4 hours). Despite these advantages, analogues of tadalafil have been rarely found, mainly due to the availability of the starting reagents, requiring piperonal that is also used in the preparation of amphetamines, whose commercialization is strictly controlled.

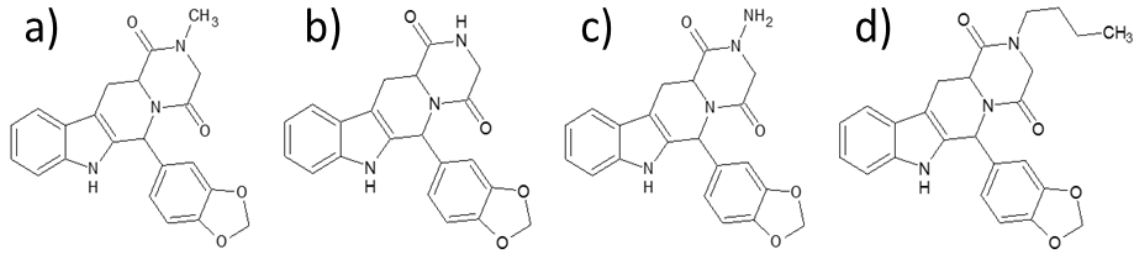

Figure 4. a) tadalafil; b) nortadalafil; c) amino-tadalafil; d) butyl-tadalafil.

The Figure 4 shows the most reported analogues of tadalafil; the major route of modification of the structure is on the N-atom of the amide. This part of the structure is non-essential and the modification do not modify the action mode.

Figure 5 reports the most common vardenafil analogues used in adulteration of food supplements, although few compounds have been reported, probably due to the fact that no important pharmacological advantages are described.

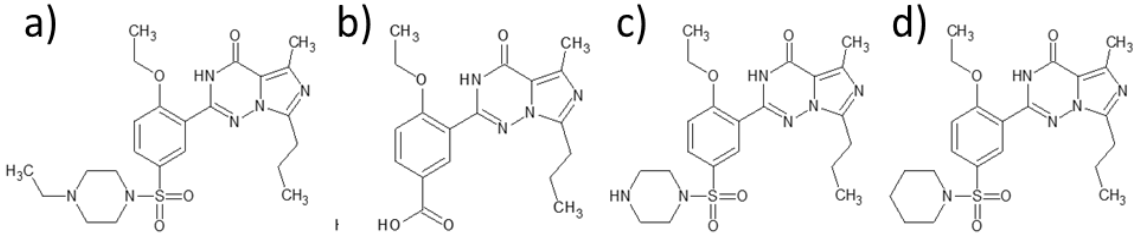

Figure 5. a) vardenafil; b) norneovardenafil, c) N-desethylvardenafil, d) pseudo-vardenafil.

Other often-used unapproved active principle is the yohimbine (Figure 6) a natural tryptamine alkaloid, which can be extracted from the bark of a variety of plants mostly of African and Asian origin such as *Pausinystalia Yohimbine*. Yohimbine hydrochloride

is rapidly absorbed and the maximum plasma concentration is generally achieved in less than one hour after oral administration.

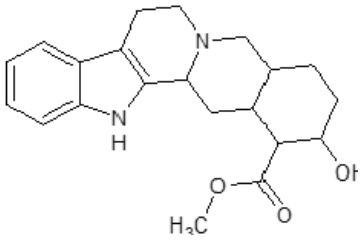

Figure 6. Yohimbine structure.

3.1.3.1 Health risk of counterfeit food supplements and normative

The possible adulteration of natural products with synthetic PDE5-Inhibitors and analogues [28-31][32-35][32-35] may be representing a risk for unaware customers [36][28][28], since PDE-5 Inhibitors show notable adverse effect such as headache, facial flushing, dyspepsia, visual disturbances and muscle pain [37][29][29]. Furthermore, analogues of the authorized PDE-5 inhibitors may have different and unknown side effect and pharmacokinetics. The health risk factor related to adulterated food supplements are mainly due to: a) the uncontrolled concentration, with cases with more than 170% of the normal dosage; b) the undefined toxicology of these compounds caused from variations of the pharmacokinetics and of the drug metabolism [26]; c) the unaware assumption of such drugs from consumers affected by incompatible sickness, as heart disease.

EU Directive 2002/46 concerning food supplements [38] allows the use of some vitamins and minerals, whilst presence of synthetic active compounds is forbidden. Therefore, any adulteration or cross contamination with PDE-5 inhibitors must be considered not in compliance with the enforced law.

3.1.4 Aerosols

The term aerosol identifies solid or liquid suspension in air, within the dimension range of 10^{-9}-10^{-4}m. Usually, in atmospheric science the term aerosols is limited only to the solid particulate suspended in the air and not to liquid, which is treated separately in the cloud science. A recent OECD report [39] forecasts that air pollution will be in the 2050 the primary environmentally cause of premature death worldwide, and identifying the increasing of the particulate matter (PM) concentration as the major contributor.

Although the link between the exposure to PM and adverse effects on human health and ecosystem is well known [40, 41], the chemistry and physic of airborne aerosol are still poorly understood; mainly because of the extreme complexity of the particulate composition and the absence of the analytical technology able to describe the whole chemistry present [42]. Aerosols are classified as emerging contaminates mainly for the increasing in the human sensitivity.

3.1.4.1 Sources, composition and size distribution of aerosols

While the 99.9% of the atmospheric content is related to the gases N_2, O_2, H_2O and Ar, molecules present at trace levels, such as NO_2, O_3 SO_2, *OH, $*O_2H$ and non-methane hydrocarbons, actually drive the atmospheric chemistry. Many different factors as gas-phase reaction, temperatures, etc. could condense low volatile species, thus forming aerosol [43, 44]. The average amount of PM in the atmosphere is in the order of 1 $\mu g/m^3$ or 10^{-7}% by mass, with a quite wide geographical and seasonal variability [45].

PM in atmosphere can be classified in many different ways taking into account different characteristic. Considering the origin of the aerosol, it can be divided in anthropogenic or biogenic. The first regards all the aerosols arising directly or indirectly from human activities and emission as for example fuel burning, while biogenic aerosols are originated from natural events as pollen, plant emission, and volcanoes emissions.

A PM mass fraction between 20-90% is represented by organic matter [46]. Organic aerosols can be emitted directly (primary organic aerosol, POA) or formed in the atmosphere through gas-to-particle conversion processes of volatile organic compounds (VOCs), thus generating what is known as secondary organic aerosol (SOA).

The aerosol composition is extremely complex because of the high chemical heterogeneity of the emissions (fuel and biomass burning) and reactions in atmosphere; a single VOC, interested by a series of complex reactions involving even oxidation by free radicals, can give thousands of different products. Furthermore, reactions inside the particles, promoted by water repartition and photochemical processes, provoke an "aging" of the aerosol with a consequent change of its composition [43].

The size of airborne matter is strongly dependent by its formation and cleaning process. There are three main modes of particulate size. The coarse mode, at high value of diameter (10 μm), is typical for the particles mechanically formed and sedimentation is the associated removal process. At the opposite in the diameter scale we find the nucleation mode, represented by ultrafine particles (<0.1μm) formed via homogenous

nucleation. These particles are generally lost by coagulation of them into the accumulation mode. Finally, the accumulation mode is described by values of diameter typically lower than 1μm (fine particles) and includes particles formed via nucleation and condensation from gas phase. Fine particles show a lifetime greater than the other two modes, and the principal removal processes are by rainout and washout.

Traditionally, in the contest of the effects of aerosol on the human health, another metric is used to describe the size of the airborne matter based on an operative concept: it is classified in PM_{10}, $PM_{2.5}$ and PM_1, referring to the particles fraction having an aerodynamic diameter lower than 10, 2.5 and 1 μm respectively.

3.1.4.2 Health and environmental effects of particulate matter

The correlation between atmospheric aerosols and human health is well established and supported by historical events and many epidemiologic studies [47]. One of the first and well documented case clearly evidencing the connection between an increased death rate and urban smog was the "great smog" occurred in 1952 in London, when an extremely high concentration of PM arising from coal burning killed an estimated number of 12,000 people [48]. More recently, episodes of PM spikes with concentrations up to 10 times greater than those reported to cause adverse health effects have been studied in Calexico/Mexicali [49]and Beijing [50][49]. $PM_{2.5}$ showed to have the heaviest impact on human health. In fact these particles can enter into the respiratory system and reach the alveoli [51]. Particles between approximately 5 and 10 μm are most likely deposited at the tracheobronchial level, while those between 1 and 5 μm are deposited at the respiratory bronchioles and the alveoli where gas exchange occurs [51].

The main human health adverse effects related to PM include premature mortality and high morbidity, asthma, cardiovascular and nervous diseases [52]. The PM toxicity shows different mechanisms: the oxidative stress is one of the major pathway affecting the respiratory system [52], involving an increased concentration of reactive oxygen species (ROS) such as superoxide radical (O_2*-), hydroxyl radical (HO*), hydrogen peroxide and other organic hydro peroxide. The ROS increment is mainly caused by the interaction with respiratory system lining fluid, but also metals and other oxygenated species contribute to this alteration.

Exposure to $PM_{2.5}$ is also correlated to the increased incidence of lung cancer, and it is estimated than 5% of lung cancer deaths are attributable to PM [53]. Carcinogenicity and mutagenicity of $PM_{2.5}$ is primarily connected to the contextual presence of certain

classes of compounds as polycyclic aromatic compounds (PAHs), coming from vehicles and biomass burning, and a number of nitrogen-containing organic compounds (NOCs) that can form carcinogenic metabolites in the body [54, 55].

In addition to the health effect, PM causes damage to ecosystems, cultural heritages, reduces visibility and it is known to be linked to climate change [40], mainly acting on the radiation balance and on absorption phenomena, warming the atmosphere or scattering radiation with a cooling effect. Even more, PM strongly affects the cloud behaviours as it acts, for its hydrophilic characteristics, as cloud condensation nuclei. [56]. PM is involved in heterogeneous reactions promoted by the solar radiation and affecting the composition of the trace compounds present in the atmosphere. Deposition of PM containing black carbon on the snow or on the surface of glacier, promotes melting phenomena by absorption of radiation, thus changing the hydrogeological cycle.

3.2 Screening analysis of emerging contaminants

In literature, many approaches for screening analysis of emerging compounds are reported. Sampling and sample preparation represent in almost all the detection approaches the preliminary stages of the analysis [9].

Several biological or biochemical methods are currently used for screening a large variety of pollutants, since interactions with the animal metabolism both at micro (enzymes and proteins) and macro (physiological effects) levels represent the main concern about the toxic effect. Historically, mouse bioassay was the first *in vivo* test implemented for pollutants detection, and it is already used in some official methods tailored for toxins analysis in food. This method not allows the identification of the compound responsible of the observed toxicity, has low sensitivity and ethical issues respect the use of animals [57].

In ELISA (Enzyme-Linked Immune Assay) assays, analytes are detected through binding to specific antibodies. Many ELISA kits are commercially available for an extremely wide variety of compounds. Despite the performance in sensitivity ELISA tests have some limitations regarding the selectivity among different variants of the same class of compounds, and the identification capacity by the fact it not provide any structural information [58].

Among the chemo-physical techniques, vibrational spectroscopy, such as infrared (IR), near infrared (NIR) [59]or Raman spectroscopy [60]or their combinations have been used for screening known contaminants, e.g. for counterfeit pharmaceutical products, and for identifying new pollutants. These techniques when used in conjunction with a chemometric approach generate typical fingerprints that help to differentiate between authentic and fake or counterfeit samples. The great advantages of this simple and non-destructive technique are limited by the reduced selectivity and sensitivity.

NMR spectroscopy is one of the most powerful tool to unambiguously elucidate the structure of known and novel compounds [61]. This technique has the advantage of an easy sample preparation and high reproducibility. The technological advances in the field of magnetic resonance have dramatically improved in terms of the sensitivity and identification capabilities by developing new NMR experiments and data processing tools. Notwithstanding NMR can be used for quantitation purpose, it often requires a quite large amount of sample, due to the low sensitivity for the trace analysis that several emerging contaminants require [61].

Mass spectrometry (MS) became increasingly common over the last decades due to its high sensitivity and selectivity. Tandem MS (MS/MS) approach allows simultaneous detection of a larger amount of analytes with increasing easier sample preparation procedure. When coupled to chromatography, MS detection offers incomparable performance for trace analysis of organic compounds. Liquid chromatography (LC), usually with a reversed phase C18 column and methanol/water or water/acetonitrile as a mobile phase, is likely the most common separation method for the analysis of polar emerging contaminants, whilst GC has been used as a separation method for volatile and semi volatile pollutants. A large number of confirmatory methods based on mass spectrometric (MS) detection were developed for the determination of contaminants in environment, food and biological matrices [62-65]. Although emerging contaminants present a large number of different compounds, for only few of them are currently available analytical standards. This fact, together with the lack of an enforced regulation for some potential pollutants, or even toxicological and structural information, have recently moved the conventional confirmatory analysis of contaminants to on non-target approach. These MS methods are based on full-scan acquisition with GC or LC coupled to high resolution (HR) MS, mainly LC-quadrupole-Time of Flight (Q-TOF) and LC-Orbitrap™ [66, 67]. LC-HRMS full scan methods do not need a compound-specific tuning, and are prone to perform non-“a priori” post-run data analysis, mainly based on

mass accuracy [68, 69]. Anyway, well-known difficulties associated with the structural elucidation, a limited availability or reliability of mass spectral libraries and software [70, 71] for non-target and post-target analysis, represent important hindrances for a widespread application of LC-HRMS techniques for identification purposes [71]. HPLC-Q-TOF [72] or HPLC-Orbitrap™ systems have been used to perform multi-residue analysis, with often more than 100 target compounds and non-target analysis.
Emerging contaminants, such as cyanotoxins can also be detected by MS without preliminary chromatographic separation. For example, MALDI-TOF instruments can be used to perform toxin analysis in very small sample volume such as cell colonies [73]. Despite the rapidity and the possibility to avoid sample preparation, this approach is not suitable for all the matrices of interest, like water, and it is poor in sensitivity and selectivity, since no information about retention times are provided to differentiate compounds with a similar MS behaviour.

3.2.1 High-resolution mass spectrometry (HRMS) in characterization of emerging contaminants.

In the paradigm of screening analysis in mass spectrometry, as represented in Figure 7, we can classify three main approaches toward the substances identification by HRMS[74]:

- **Target analysis**: This approach is focused on the confirmation and quantitation of a limited number of compounds. The number of compounds determinable depends on the specific detection system and experimental design. Target screening requests the use of analytical standards and mass spectrometry has to be coupled with a chromatographic separation step.
- **Suspect screening**: this approach is performed on a relatively large number of selected compounds, whose presence in the sample is supposed, even if the corresponding certified standards are not necessary available. Usually, information about structures of the suspected compounds is included in a database in order to simplify data analysis. Through this screening approach is possible to reach identification or confirmation of substances when certified standards are available.
- **Non-target screening**: this approach is followed when no structural information is available *a priori* and the analysis is virtually performed on all substances detectable in the sample analysed.

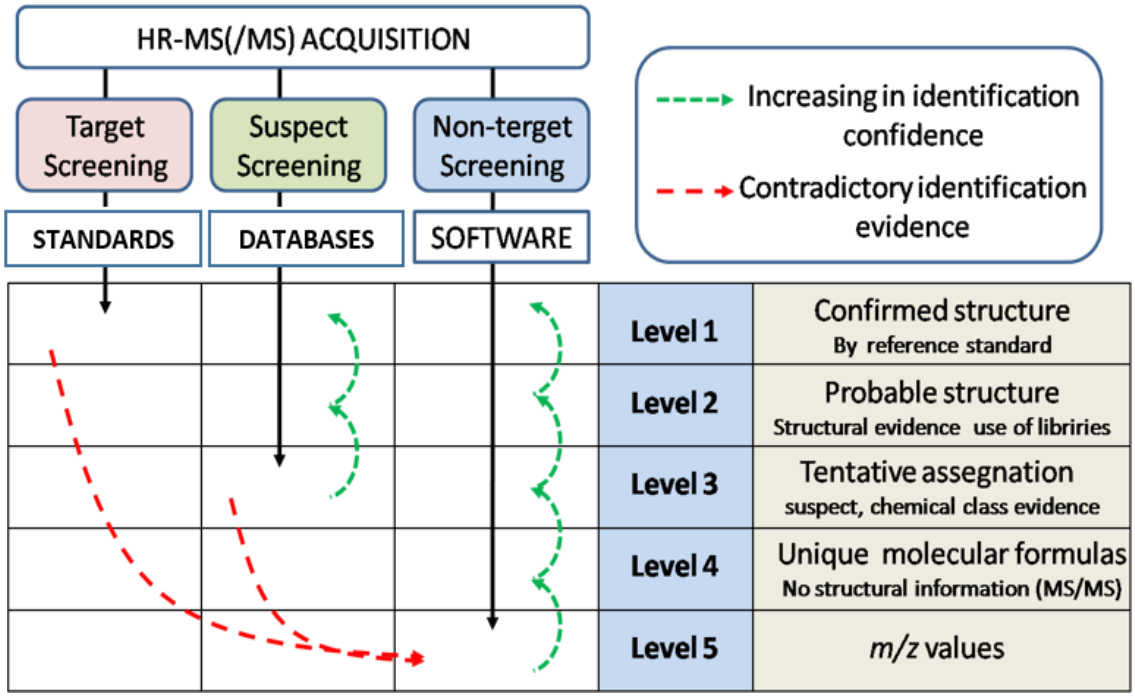

Figure 7. Representative scheme of screening analysis and correlation between acquisition approach and level of identification provided.

The use of standards, specific databases and software utilities are necessary depending on the screening approach followed.

3.2.2 Degree of identification

A consideration regarding the harmonization of the definitions in "identification" and "confirmation" has to be done, because the scientific community often uses them improperly. When acceptable evidences support the structure of an analyte, we can use the term "identification", while a comparison with an analytical standard through orthogonal method has to be provided in order to have "confirmation" [75].

In literature, and generally in the field of environmental analysis, the identification of analytes by LC-MS mainly refers on the European guideline for the confirmation of veterinary drugs in food of animal origin, the EU Guideline 2002/657/EC [76]. In this guideline, the concept of identification points (IPs) is proposed for confirmatory methods that anyway require a comparison with analytical standards. The identification of a target analyte has to comply with the following constrictions:

- **Reference standard matching**: analytical standards should be analysed contextually to the sample analysis. All signals and parameters related to the suspected compounds (e.g. retention time and mass spectra) have to match with

those obtained for the analytical standard in the same experimental conditions, with acceptable bias. The use of reference materials is encouraged, and if not available, the use of spiked blank matrix is desirable instead of a pure standard. When the matrix affects performance of the separation or detection system, the use of pure standard has to be avoided.

- **Chromatography**: the chromatographic peak of a positive sample should exceed a signal to noise ratio (S/N) threshold of 3:1, and the retention time have to match with the peak of the corresponding standard with a maximum relative standard deviation of 2.5% for LC and 0.5% for gas chromatography.
- **Mass Spectral matching**: the mass spectrum should include a number of signals satisfying the IPs requested, at least three. The number of mass signals requested depends on the MS system used, basically on low or high resolution mass analysers, and on single stage or tandem MS. Thus, the MS fingerprint has to match that obtained for the reference standard, also satisfying specifications on the S/N, and relative intensities.

In the framework of confirmatory analysis, the chromatographic separation is absolutely necessary for identification purpose, because it indirectly provides further information on substance structures, thus allowing a unique identification when standards are available. However, the retention time parameter could be useful even in absence of certified standards, if a suitable normalization is adopted. As example, retention time index (RTI) or Kovats index could be used as further descriptor for databases in order to improve the reliability of an identification process.

Finally, when HRMS systems are employed, a separation system could be not necessary, if the aim of the analytical protocol is a qualitative characterization of the organic composition based on the raw formulas inferred by experimental measures.

Alongside the official guidelines, different approaches of classification have been proposed in the literature. Recently, the concept of "identification level" has been widely discussed in the environmental scientific community. One proposal originates from the field of metabolomics, and is strongly supported by many European institutions, and among them the NORMAN association in which our research group was involved [74]. According to this approach, and relying on information provided by the non-target analysis, we can define 5 levels of identification confidence as show in the Figure 7:

Level 1. Confirmed structure. This level of identification confidence corresponds to the confirmatory analysis previously described. In order to provide a unique structural identification, orthogonal selective methods have to be used, and data obtained have to be compared with the corresponding ones obtained from analysis of certified standards. Mass spectrometry has to be necessarily coupled with a chromatographic separation step.

Level 2. Probable structure. An adequate amount of evidences, comprising MS/MS diagnostic fragments, indicates a unique structure of the analyte. The main difference with the level 1 is the absence of the comparison with a standard. This degree of confidence can arise from matching spectra with MS libraries or literature.

Level 3. Tentative candidates. In this level of confidence, structural information is significant but not sufficient to ensure a unique structure. MS/MS fragmentation is available, giving some indications for example, of the chemical class of compounds.

Level 4. Unequivocal molecular formula. In this case, a molecular formula can be unambiguously assigned by the MS data. No significant MS/MS spectra are recorded.

Level 5 Mass of interest. It is the lowest level of identification confidence, and it is limited to the knowledge of the *m/z* of the quasimolecular ion. In level 5 all the compounds are classified as "unknown".

For both suspect screening and non-target analysis, it is possible reach the highest level of identification, gaining informationally. Incidentally, while the use of chromatography is mandatory for the confirmation, it is not necessary in the case of the lower level of identification.

3.2.3 Identification criteria in HRMS

3.2.3.1 Molecular and quasimolecular ion

When soft ionization are employed, molecular or quasimolecular ion is considered the MS signal giving the most important information in target and suspect screening analysis, as from its *m/z* value is possible the direct determination of the molecular formula. Molecular or quasimolecular ions have to be optimized in term of sensitivity and trueness. The last parameter results pivotal for the determination of unknown compounds, when HRMS is employed for screening purpose. Two parameters characterize the measure of the MS trueness: resolving power and mass accuracy. Resolving power or resolution (R) is the ability of a mass analyser to resolve

neighbouring signal. It is defined as m/Δm where m is the nominal mass of a given molecule and Δm is the difference in mass that the instrument is able to discriminate. This definition is nowadays subjects to many critical revisions in particular about how it is defined. An alternative definition, used even in the 2002/657/EC guideline is the full width at half maximum (HWHM).

The instrumental mass accuracy refers to the degree of closeness of a measured *m/z* to its true theoretical value. Accuracy better than 2 ppm and 5 ppm are nowadays quite common for the new generations of Orbitrap™ and Q-TOF mass analysers respectively and sub-ppm accuracy could be archive by FTICR.

The accurate mass of both molecular (and/or quasimolecular) and fragment ions is the parameter used in the libraries and online database in order to find the corresponding structure of possible unknown compounds. In the Table 1 the main libraries and databases available in MS are reported. While NIST, PubChem, ChemSpider and Wiley libraries can be considered general libraries of greater value, other databases as NIST MS2, MassBank, METLIN, mzCloud, are becoming powerful tools for identification in HRMS because a quite large amount of MS/MS spectra is currently available, with good perspectives of improvements.

Table 1. List of the most used in screening analysis by mass spectrometry databases and libraries and MS spectra included.

Database or library name	total compounds present	Entries with MS spectra
ChemSpider	32000000	-
DAIOS	1404	>1000
PubChem	68479719	-
STOFF-IDENT	7864	-
MassBank MS/MS	3350	3350
mzCloud	2510	1956
NIST EI-MS	242477	212961
NIST MS/MS	8171	4628
Wiley Registry of Mass Spectral Data (EI-MS)	638000	289000
Agilent Broecker, Herre & Pragst Toxicology/Forensicsf	8998	
Agilent METLIN Pesticide Library	1664	
Agilent METLIN Synthetic Substance Libraryg	64092	
Agilent METLIN Veterinary Drug Library	1684	
Bruker ToxScreener (incl. Pesticide Screener)	1753	
Sciex / AB Sciex LC/MS/MS Meta Library	2381	
Thermo Environmental Food Safety (EFS) with/without retention	732	
Thermo toxicology	654	
Waters database	730	

3.2.3.2 Isotopic pattern

The isotopic abundance can provide additional information on elemental formulas. The relative abundances of the various isotopomers are helpful for prioritizing and reducing the number of the possible formulas assigned, and this approach is implemented on many software tools for MS data elaboration, giving a weight at the final score on the basis of the isotope ratios and mass defects of isotopes [77].

Due to the low intensity of the isotopomer ions, high sensitivity and a low noise with few chemical and background interferences are required. Obviously, some chemical elements show a peculiar isotopic abundance, e.g. chlorine, bromine, silicon sulphur and several metals. Thus, such MS spectra result in a typical fingerprint that allows

some consideration about the presence of certain element, empowering a possible identification.

3.2.3.3 Fragmentation

MS/MS is a powerful technique ensuring information that increase the reliability of the molecular formula assignment and provide structural information for the identification. Official guidelines assign IPs for each fragment present matching with the fragmentation pattern of the corresponding pure standard. In addition to this approach, when commercial standards are not available, the matching with MS/MS spectra collected in libraries can be considered to strengthen the identification.

Instrumental approaches for acquiring fragmentation spectra can be different and dependent on the apparatus available and on the aim of the analysis performed. A recent trend in non-target analysis is the collection of MS/MS spectra of as many analytes as possible using data-dependent acquisition features. In other instrumental approach, the so-called MS^E ™ involving the simultaneous acquisition of accurate mass data at low and high collision energy can provide this information.

In silico fragmentation is an alternative recent approach for structure identification. Mass Frontier and ACD/MS Fragmenter are the two most popular MS fragmentation predictor's rule-based programs. They are useful to predict MS/MS spectra of compounds when no reference mass spectra are present. Despite the potential capabilities, in-silico fragmentation has been proved to be efficient for only 56% of mass peaks, and often fails with analytes at trace level in noisy spectra. However, in-silico fragmentation remains the principal approach in protein identification.

3.3 HRMS data visualization tools

When HRMS is used in non-target analysis generates an extremely large amount of data as list of assigned molecular formulas [78]. The interpretation of any single compound is extremely time consuming and useless. Some commercial software's provide tools for statistical evaluation of data, using all the possible parameters, like retention time, molecular and fragment ions, signal abundance, for blank subtraction and classification of results.

In the framework of the categorization of aerosol visualization methods are generally necessary for data evaluation. The most diffused visualization methods include the use of double bond equivalent (DBE), van Krevelen diagrams, carbon oxidation state and Kendrick mass analysis [79, 80].

DBE is an indicator of the hydrogen deficiency and represent the number of the double bonds and rings in a molecule structure [79]. For molecules with general formula CcHhOoNn, DBE can be calculated through the general formula:

$$DBE = c - 0.5h + 0.5n + 1$$

Where c, h and n are the number of atoms of carbon, hydrogen and nitrogen respectively. This value is a useful tool to eliminate the molecular formula showing an unreasonable high number of unsaturations and high double rings. Usually, the data are visualized as DBE against number of carbon or *m/z* ratio. Anthropogenic emission are characterized by a high grade of unsaturations and values of DBE (>5) typical of aromatic hydrocarbons and their oxidized derives. DBE plots can provide useful information about the sources and precursors of aerosols.

Van Krevelen (VK) diagrams were initially developed for the study of coalification processes and then extended to the particulate matter [81]. In these diagrams, H/C ratio of formulas are plotted versus O/C ratios. VK diagrams are particular useful to classify aerosol samples by identification of the compound classes present on it. Different classes of compounds occupy different regions of the plot. Highly unsaturated compound (PAHs and derivatives) have low H/C and O/C ratios and lie close to the axes origin; aliphatic hydrocarbons (lipids as example) have again low O/C ratio but high H/C ratio and occupy a region above the previous ones. Highly functionalized compound are characterized by a high value of O/C ratio and the typical region of the

plot for those compound is the right side. Moreover, those plots are often used in three-dimension with signal intensity or the DBE value in the z-axis in order to maximize the information. Anyway, signal intensity is a parameter that has to be carefully evaluated when the direct infusion of sample is used for the analysis, as the direct correlation with compounds concentration is not possible.

Another useful parameter to characterize the extreme chemical complexity of organic compounds present in atmospheric aerosols is the oxidation state of carbons (OSc) Oxidative reaction plays a central role in the atmospheric chemistry and are involved in the removal of pollutants, O_3 formation and SOA production. In fact, as reported in literature SOA is mainly formed by organic compound arising from the oxidation of gas-phase species. Thus, the oxidation state of carbons could be a metric to characterize SOA, studying its formation and temporal evolution. The trend in the atmospheric environment is the increasing of the oxidation state of carbon, through bonds formation with oxygen and breakage of hydrogen carbon bonds. OSc can be calculated taking in to account the oxidation state of each heteroatom present in the structure and applying the follow formula:

$$OSc = -\sum_i OS_i \frac{n_i}{n_c}$$

Where OS_i is the oxidation state of i-element and ni/nc is the molar ratio of element i to carbon. The formula is accurate when the structure of the compound is exactly known [78, 82]. When it is applied on molecular formula it could be inaccurate if in the formula are present elements showing different oxidation states possible depending of the molecular structure. As example nitrogen in organic compound can be present as N (-3) in ammines, N (-1) in amine oxide compounds, N (+1) in nitrous-compounds and N (+3) in nitro-compounds. A similar variability is also showed by sulphur that can be present in the oxidation states -2 (thiols and sulphides), 0 (sulfoxides), +2 (sulfinic acid compounds), +4 (sulfonic acid compounds). This variability in oxidation states make difficult to predict OSc from formulas containing nitrogen and sulphur and its use should be avoided. In this study, OSc is calculated only for the species CcHhOo using the simplified formula [78]:

$$OSc = \frac{n_h}{n_c} - 2\frac{n_o}{n_c}$$

Despite the fact that oxygen shows multiple valence state, the (-2) is much more stable than the others and then the sporadic presence of formulas containing oxygen with different oxidation state does not significatively affect the overall analysis.

Kendrick mass defect (KMD) analysis is based on the fact that different nucleotides show different defects of mass from the nearest integer mass [83]. Therefore, different elemental compositions showing same integer mass have different exact mass. The addition or subtraction of two hydrogen atoms from a molecule means increase or decrease the number of unsaturations present on it by 1. In the same way addiction of a CH_2 group, do not affect the unsaturation number but involve in a shift of mass and defect of mass (14.01565). In KMD analysis the masses of molecular formulas CcHhOoSsNn are rescaled converting the exact mass of CH_2 to the nominal one applying the formula

$$KM = IUPAC\ mass\ \times (14/14.01565)$$

In these condition formulas, differing only by the number of CH2 will have identical mass defects. The KMD could be calculated with the formula:

$$KMD = nominal\ mass\ -\ Kendrick\ mass$$

In the visualization usually KMD are plotted on y-scale versus Kendrick mass on x-scale. Molecular formulas differing by number of CH_2 are displayed in horizontal series spaced by 14 Kendrick mass units. Instead, different homologue series are plotted at different KMD values and compound classes differing by 2H fall in lines spaced by 2.01564 * (14/14.01565) - 2 = 0.01340. KMD analysis is particularly useful to identify classes of compounds forming series differing by number of CH_2 [83].

4 Aim of the project and structure of the thesis

In this PhD thesis, the capability of HRMS in screening analysis will be investigated for the study of emerging contaminants in environmental matrices and foodstuff. In details, we would focus the attention on the process of scale-up in the identification confidence by developing of different specific protocols.

Advantages in the screening analysis must be driven by the availability of analytical standards and the implementation of useful databases and software. These features will be treated taking into account the identification level we want obtain. Chapter "Development of a target and suspect two-step protocol for analysis of freshwater cyanotoxins by liquid chromatography-Q-TOF system" and chapter "and chapter **"**One-shot" analysis of PDE-5 inhibitors and analogues in natural products for the treatment of erectile dysfunction" will treat the highest levels of identification confidence (level 1 and level 2) and reports the developed methods to perform simultaneously the suspect and target screening.

Chapter "HRMS analysis of organic fraction in $PM_{2.5}$: Post-run data analysis workflow and the role of ionization sources" will treating the low identification confidence levels (level 5, level 4) with emphasis on the crucial step of molecular formula assignment to *m/z* values.

Here below, the aims of the single chapters are introduced.

4.1 Development of a two-step method for target and suspect analysis of freshwater cyanotoxins by LC/Q-TOF system.

Occurrence of cyanotoxins in waters intended for human consumption and foodstuff requires a reliable analytical strategy able to support rapid decisions. Moreover, health agencies usually have to opt for a limitation of the drinking water distribution, without a clear legal orientation or a well-established risk assessment about the presence of MCs variants other than MC-LR, or other potentially toxic oligopeptides. Thus, specific

confirmatory analysis aimed to furnish as many information as possible on cyanotoxins risk management are valuable.

To our best knowledge, nor rational LC/HRMS-based analytical protocols neither specific databases devoted to determine standardless cyanotoxins in freshwater have been reported. Thus, our effort was focused in developing a reliable and rapid strategy, useful to risk assessment related to cyanotoxins.

4.2 Development of "one-shot" analysis of PDE-5 inhibitors and analogues in natural products for the treatment of erectile dysfunction.

This study started form a collaboration between our research group, the Italian agency for drugs (agenzia Italiana del farmaco, AIFA), the Ministry of Health, the National Institute of Health (Istituto Superiore della Sanità; ISS) and the Department of Scientific investigations of the Police. The aim of the collaboration was the monitoring of commercial food supplements sold for the treatment of erectile dysfunction and the verification of the compliance of these products with the legal requirements. For this purpose, an analytical method for determining the presence of PDE-5 inhibitors and analogues able to perform simultaneously target and suspect screening of compounds with such pharmaceutical potential was needed. Seven compounds (yohimbine, sildenafil, vardenafil, tadanafil, homosildenafil, pseudovardenafil and hydroxyhomovardenafil) have been selected among PDE-5 inhibitors for the target analysis.

4.3 Development of a workflow for HRMS analysis of $PM_{2.5}$ organic fraction: post-run data analysis and the role of ionization sources.

These activities were mainly carried out in the frame of six months of research conducted during the PhD at the Cambridge University, under the supervision of prof. Markus Kalberer. A method of analysis of organic fraction of $PM_{2.5}$ based on a methanol extraction of samples and nanoESI-MS determination was previously

developed in the Kalberer's group [84]. Two different algorithms were used for the data elaboration by using nanoESI in positive or negative acquisition mode. However, these protocols were time-consuming and affected by the presence of false positive results.

The purpose of the research activities was twofold:

- the evaluation of the Atmospheric Pressure Photoionization (APPI) source for performing analysis of the $PM_{2.5}$ organic fraction, which could be well ionized using this technique, since it is known to have highly unsaturated compounds, like polycyclic aromatic hydrocarbons (PAHs), nitro and oxidized PAHs.
- the optimization of the algorithms used for the data treatment in order to consider different signals arising from the ionization sources, and to reduce both false positive and the elaboration time.

5 Materials and Methods

5.1 Reagents and samples

MC-RR, MC-YR, MC-LR, MC-LA, MC-LW, MC-LF, MC-LY, MC-HtyR, MC-HilR, MC-WR [D-Asp3]-MC-RR, [D-Asp3]-MC-LR , and nodularin (NOD), used as internal standard (IS), were purchased from Alexis® Biochemicals.

Stock solutions of the twelve MCs and IS, were prepared by dissolving each compound with at least 2 mL of methanol. Subsequent dilutions to obtain working standard solutions were obtained by suitable diluting stock solutions with mobile phases. All solvents and chemicals were of analytical grade (Sigma Aldrich) and all standard solutions and water samples were stored at -18°C in the dark to minimize analyte degradation.

Analytical standards (HPLC grade) of Yohimbine, sildenafil, Vardenafil, Tadalafil, Homosildenafil, Pseudovardenafil and Hydroxyhomovardenafil were purchased by LGC Standards . Formic acid (FA, reagent grade, min 99%), trifluoroacetic acid (TFA, reagent grade, min 99%), pentafluoropropionic acid (PFPA, reagent grade, min 99%), acetonitrile and methanol, both of LC-MS grade, were purchased from Sigma-Aldrich. Water was purified using a Milli-Q Water System (Millipore) to 18 MΩcm. Homosildenafil, Vardenafil, Hydroxyhomosildenafil, Pseudovardenafil e Tadalafil stock solutions were prepared at concentration of 1.0 mg/mL in acetonitrile 0.1% FA, while Yohimbine e sildenafil stock solutions were prepared at the same concentration in a water: acetonitrile 50:50 (v/v) 0.1% FA solution.

LC-MS RTI mixture was prepared by the Technische Universitat Munchen – Ingenieurfakultat- Bau Geo Umwelt – Lehrstuhl fur Siedlungswasserwirtschaft. The concentration of each analyte (

Table 2) in the mixture is approximately 10µM dissolved in ACN:H_2O 50:50 (v/v).The solution was stored at 4 °C in the fridge.

A standard mixture of endocrine disruptors containing the compounds reported in the Table 3 was used in the confirmation of some positive target compounds in NORMAN trial. The stock solutions were in methanol:H_2O 50:50 and stored al-18°C. Further dilutions were obtained using the same solvent.

Table 2. RTI standard mixture constituents.

CAS	Standard	monoisotopic MW [Da]	logP
657-24-9	Metformin	129.1014	-1.36
1698-60-8	Chloridazon	221.0355	1.11
16118-49-3	Carbetamide	236.1160	1.65
150-68-5	Monuron	198.0559	1.93
3060-89-7	Metobromuron	258.0004	2.24
13360-45-7	Chlorbromuron	291.9614	2.85
125116-23-6	Metconazole	319.1451	3.59
333-41-5	Diazinon	304.1010	4.19
124495-18-7	Quinoxyfen	306.9966	4.98
49562-28-9	Fenofibrate	360.1128	5.28

Table 3. Endocrine disruptors' analytical standard and stock solution concentration.

Compound name	Abbreviation	molecular formula	$[M-H]^-$ *m/z*	Stock solution (mg/L)
4-N-octylphenol	n-OP	$C_{14}H_{22}O$	205.1598	1000
4-tert-Octylphenol	t-OP	$C_{14}H_{22}O$	205.1598	1000
Nonylphenol	NP	$C_{15}H_{24}O$	219.1754	1000
Bisphenol A	BPA	$C_{15}H_{16}O_2$	227.1078	1000
Bisphenol A deuterate	BPA-D	$C_{15}D_{16}O_2$	243.1329	1000
Perfluoro-n-pentanoic acid	PFPeA	$C_5HF_9O_2$	262.976	1000
Estrone	E	$C_{18}H_{22}O_2$	269.1547	1000
β-Estradiol	b-Eol	$C_{18}H_{24}O_2$	271.1704	1000
β-Estradiol deuterate	b-Eol-d	$C_{18}D_3H_{21}O_2$	274.1852	1000
17α-ethinyl estradiol	a-Etinil E	$C_{20}H_{24}O_2$	295.1704	1000
Perfluorobutanesulfonic acid	PFBS	$C_4HF_9O_3S$	298.943	1000
Perfluorohexanoic Acid	PFHxA	$C_6HF_{11}O2$	312.9728	1000
Perfluoroheptanoic Acid	PFHpA	$C_7HF_{13}O_2$	362.9696	1000
perfluorohexane sulfonate	PFHxS	$C_6F_{13}O_3S$	398.9366	1000
Perfluorooctanoic acid	PFOA	$C_8HF_{15}O_2$	412.9664	1000
Perfluoro-n-decanoic acid	PFNA	$C_9HF_{17}O_2$	462.9632	1000
Perfluorooctanesulfonic acid	PFOS	$C_8HF_{17}O_3S$	498.9302	1000
Perfluoro-n-decanoic acid	PFDA	$C_{10}HF_{19}O_2$	512.96	1000
Perfluoro-n-undecanoic acid	PFUnDA	$C_{11}HF_{21}O_2$	562.9569	1000
Perfluoro-n-dodecanoic acid	PFDodA	$C_{12}HF_{23}O_2$	612.9537	1000

A stock standard mixture of PAHs, Nitro-PAHs and oxidized PAHs (O-PAHs) (Supelco, grade TraceCERT®) was diluted in methanol:dichloromethane 1:1 to obtain the diluted standard solution reported in the Table 4. The concentrations were in the range 6-133 µg/mL for PAHs, 0.6-5.3 µg/mL for Nitro-PAHs and 0.13-13 µg/mL for O-PAHs. The solution was stored at -18°C to prevent degradation.

Table 4. PAHs, oxo-PAHs and nitro-PAHs standard solution composition.

Compound	Molecular formula	Stock solution (µg/mL)	Diluted St. MIX (µg/mL)
Acenaphthene	$C_{12}H_{10}$	1000	66.67
Acenaphthylene	$C_{12}H_{8}$	2000	133.33
Anthracene	$C_{14}H_{10}$	100	6.67
Benz[a]anthracene	$C_{18}H_{12}$	100	6.67
Benzo[b]fluoranthene	$C_{20}H_{12}$	100	6.67
Benzo[k]fluoranthene	$C_{20}H_{12}$	100	6.67
Benzo[ghi]perylene	$C_{22}H_{12}$	200	13.33
Benzo[a]pyrene	$C_{20}H_{12}$	100	6.67
Chrysene	$C_{18}H_{12}$	100	6.67
Dibenz[a,h]anthracene	$C_{22}H_{14}$	200	13.33
Fluoranthene	$C_{16}H_{10}$	100	6.67
Fluorene	$C_{13}H_{10}$	200	13.33
Indeno[1,2,3-cd]pyrene	$C_{22}H_{12}$	100	6.67
1-Methylnaphthalene	$C_{11}H_{10}$	1000	66.67
2-Methylnaphthalene	$C_{11}H_{10}$	1000	66.67
Naphthalene	$C_{10}H_{8}$	1000	66.67
Phenanthrene	$C_{14}H_{10}$	100	6.67
Pyrene	C16H10	100	6.67
9-nitroanthracene	$C_{14}H_{9}NO_{2}$	1000	0.67
4-nitrocatechol	$C_{6}H_{5}NO_{4}$	1000	5.33
4-nitrophenol	$C_{6}H_{5}NO_{3}$	1000	2.67
9,10-antraquinone	$C_{14}H_{8}O_{2}$	1000*	13.33
9-phenanthrenecarboxaldehyde	$C_{15}H_{10}O$	1000	2.67
9-fluorenone	$C_{13}H_{8}O$	1000	2.67
1-naphthaldeyde	$C_{11}H_{8}O$	1000	5.33
9-hydroxyphenanthrene	$C_{14}H_{10}O$	500	0.13
9-hydroxyfluorene	$C_{13}H_{10}O$	500	0.13

5.2 Samples and sample preparation

5.2.1 Freshwater samples

All water samples (27) analysed for cyanotoxins determination were selected among freshwaters intended for human consumption or drinking water affected by cyanobacterial blooms. Samples have been collected and processed by the Department of Inland Water of the National Institute of Health (ISS), accordingly to the analytical procedures described in a previous research [12]. The procedure consist in a preliminary freezing of the sample; this step directly damage cell membranes and release intracellular toxins. Cell lysis prior to filtration allows the simultaneous detection of both extracellular and intracellular toxins. The sample concentration was carried out by extraction and clean-up through a Graphitized Carbon Black cartridge and elution with organic solvent.

A list of samples, together with the cyanobacteria identified during the morphological analysis is reported in Table S 1. Twenty µL of the sample extracts were injected onto the LC/MS system. Conversely, the water sample named Bidighinzu, was injected directly into the detection system (40 µL), since the concentration of cyanotoxins was expected to be quite large.

5.2.2 Food supplements.

The following commercial pharmaceutical formulations were purchased in drugstores: Cialis® from EliLilly (Sesto Fiorentino , Italy), Levitra® from Bayer Pharma AG (Berlin, Germany) and sildenafil from DOC Generici (Milan, Italy). Two sets of food supplements and herbal dietary products were collected in this project. The first one, consisting in 21 bulk material for herbal products were supplied by a local seller, while the second set for a total of 5 samples were collected during a market monitoring campaign coordinated by the Italian Medicine Agency (AIFA) and in collaboration with the special anti-adulteration police force (N.A.S). These samples were bought in various shops (sexy-shops, herbalist's shop and ethnic shops) in four Italian cities.

Figure 8. Examples of food dietary supplement analysed.

After grinding solid samples in a ceramic mortar, about 10 mg of each matrix were weighted into an Eppendorf tube. Each sample was extracted with 1.5 mL of water: acetonitrile 50:50 (v/v) acidified with 0.1% FA for 10' in a sonic bath. After centrifugation at 10000 RPM for 10' (Mikro120, Hettich), an aliquot of 0.5 mL of the supernatant was transferred in a 1.5 mL vial for auto-sampler. For pharmaceutical formulations, the extracts were diluted by a factor 1000 to bring the analyte concentrations inside the range of the calibration curve.

5.2.3 NORMAN trial sample.

The sample used in the collaborative trial was collected from location JDS57 on the Danube River, downstream of Ruse/Giurgiu on the 18th of September 2013. The sample had been prepared by large-volume solid-phase extraction (LVSPE) of 1000 litres of river water . The sampler cartridge was composed by 160 g of Macherey Nagel Chromabond® HR-X (neutral resin), 100 g each of Chromabond® HR-XAW (anionic) and HR-XCW (cationic exchange resin). The resins were extracted with 500 mL each of ethyl acetate and methanol (HR-X), 500 mL methanol with 2% of 7 M ammonia in methanol (HR-XAW) or 500 mL methanol with 1% formic acid (HR-XCW). The extracts were then combined, neutralized, filtered (Whatman GF/F) and reduced to a final volume of 1 L using rotary evaporation. Aliquots of 1.5 mL, equivalent to 1.5 L of river water, were transferred into vials and evaporated to dryness under nitrogen. The sample war reconstructed with 1.0mL of H_2O:AcN 50:50, and an aliquot of 40µL was injected into the LC-MS apparatus**.**

5.2.4 Urban $PM_{2.5}$

Ten Teflon filters (47 mm, 0.2 µm) were pre-treated for removing organic contaminants. Filters were washed for 30 minutes in ultrasound bath successively with 2 X 20mL of deionized water, 2 X 20mL of acetonitrile and 2 X 20 mLof methanol. Finally, filters were dried under vacuum for one hour and stored in a clean desiccator.

Six samples of $PM_{2.5}$ were collected for 24 hours in the sampling room of the Department of Chemical Sciences of Padua (45.41°N, 11.88°E), while four filters were left unused as procedural blanks. A Zambelli Explorer Plus PM sampler equipped with proper inertial impactors was placed in a sampling room equipped with a wall fan which sucks air from outside making the room representative of Padua urban background air. The sampling system was fitted with $PM_{2.5}$ certified selectors (in 2006, CEN standard methods UNI-EN 12341 and UNI-EN 14907) working at a constant flow rate of 38.3 L/min (2.3 m^3/h) and equipped with Ø 47 mm Teflon (PALL, Teflon Membrane, 1 µm pore size).

Further six samples, collected on quartz filters in Mandria, a suburban area of Padua, were obtained from ARPAV (Regional Agency for the environmental protection), together with their respective blank samples. Table 5 reports the specification of the samples collected.

Table 5: Sampling specification.

Date	Filter's code	Volume (L)	PM 10 Conc. (µg/m3)	Rain precipitation (mm)	Average Temperature (°C)	Filter type
08/01/2015	FP1-080115	54947	96	0	11.1	PTFE
09/01/2015	FP2-090115	55022	113	0	10.2	PTFE
10/01/2015	FP3-100115	55192	113	0	10.0	PTFE
12/01/2015	FP4-120115	54943	87	0	10.5	PTFE
13/01/2015	FP5-130115	55047	80	0	10.9	PTFE
14/01/2015	FP6-140115	55051	91	0	10.7	PTFE
02/06/2014	Q2-020614	N/A	16	0	19	Quartz
04/06/2014	Q3-040614	N/A	17	8	19	Quartz
06/06/2014	Q4-060614	N/A	33	0	22	Quartz
09/06/2014	Q5-090614	N/A	34	0	26	Quartz
17/06/2014	Q6-170614	N/A	11	0	21	Quartz
19/06/2014	Q7-190214	N/A	17	0	21	Quartz

All glassware and taps used in the extraction procedure were accurately cleaned using at least three washing with HPLC grade methanol (Fluka).

A quarter of filter was manually cut and the portions of filter on the borders not containing sample were removed. The samples were extracted three times either with 5mL of methanol in ultrasound bath at 0°C for 30 minutes. Low temperature is needed to avoid the methylation of carboxylic groups by reaction with methanol. The extracts were combined and filtered through PTFE filter 0.45µm and 0.22µm and then evaporated at 29°C under gentle nitrogen steam until the final volume of 1.0 mL.

5.3 Instrumental analysis

5.3.1 Cyanotoxins in drinking water

LC-Q-TOF-MS analysis were performed with an ultra-high pressure liquid chromatography (UHPLC) system (Agilent Series 1200; Agilent Technologies, Palo Alto, CA, USA), consisting of vacuum degasser, auto-sampler, a binary pump and a column oven coupled to both Diode-Array Detection and Q-TOF-MS mass analyser (Agilent Series 6520; Agilent Technologies, Palo Alto, CA, USA).

The analytical column was a Kinetex C-18 (2.6µm 100 mm x 2.1 mm i.d., Phenomenex, Italy) and it was thermostated at 30°C. The sample-injected volume was 20 µL. The mobile phase components A and B were water and acetonitrile respectively, both acidified with 10 mM formic acid. The eluent flow rate was 0.3 mL/min. The mobile phase gradient profile was as follow (t in min): t_0, B= 20%; t_{10}, B= 55%; t_{11}, B= 80%; t_{15}, B= 100%; t_{19}, B= 100%; t_{20}, B=20%; t_{25}, B=20%.

The Q-TOF system was equipped with an ESI, operating in dual ESI mode and positive ESI acquisition, with the following operation parameters: capillary voltage, 3500 V; nebulizer pressure, 35 psi; drying gas, 8 L/min; gas temperature, 350°C; fragmentor voltage, 180V; skimmer 65 V.

5.3.1.1 Two-step confirmation analysis and suspect screening protocol

This protocol based on HPLC-Q-TOF platform, consists in a preliminary analysis in full scan MS mode followed by a data processing step involving a customized library for detecting the MS signals of interest. A further acquisition analysis in target MS/MS mode provided the structural information needed for the identification of the cyanotoxin

variants. The use of standards in both the analysis mode allowed the confirmation of the analytes.

Full scan mass spectrum was recorded as centroid over the range 50–2000 *m/z* with a scan rate of 2 spectra/s and analysed by the software Masshunter to find the suspect analytes. A Metlin database was generated and used within the Molecular Features Extraction (MFE), setting the following parameters and thresholds: a) ion compound filters ≥ 1000 in MS level, b) retention time ≤ 2.5% of tolerance with respect to selected cyanotoxins standards and c) MME ≤ 20 ppm.

Target MS/MS analysis were performed at 8 spectra/s over the *m/z* range 50-2000 and CID energy of 45eV for quasimolecular ions and 20eV for double-charged species.

The Q-TOF calibration was daily performed by using the manufacturer's solution. For all chromatographic runs, the *m/z* 391.2843 relative to the diisooctylphthalate molecular ion, always present as impurity, was set as lock mass for accurate mass analysis. The instrument provided a typical resolving power (FWHM) of about 18000 at *m/z* 311.0805.

5.3.1.2 Auto MS analysis and suspect screening protocol.

This protocol consists in recording in the same chromatographic run the full scan and MS/MS spectra of the suspect analytes by the "autoMS/MS" feature of the mass analyser.

The same database used in the two-step protocol described above, was converted into a .csv file to be used as a preference list for auto MS scan acquisition mode.

The same HPLC conditions and source parameters as described above were used in this screening approach. In this MS feature the following parameters were set: MS and MS/MS scan speed 6 and 8 spectra/s respectively, for a total cycle of 1.26 s, 8 compounds per cycle, isolation width of 4 *m/z*, active mass exclusion enabled after 100 spectra and 0.3 min, absolute and relative precursor threshold 200 counts and 0.001 % respectively.

5.3.2 PDE-5 inhibitors in food supplements

Q-TOF-MS analysis were performed with an HPLC system (Agilent Series 1200; Agilent Technologies), consisting of vacuum degasser, auto-sampler, a binary pump and

a column oven coupled to a Q-TOF-MS mass analyser (Agilent Series 6520; Agilent Technologies).

Separation was achieved with a PolymerX™ (5 µm, 150 mm x 2 mm i.d., Phenomenex, Italy) thermostated at 30°C. The mobile phase components A and B were water and acetonitrile, respectively, both acidified with 0.1% of FA. The eluent flow rate was 0.3 mL/min. The mobile phase gradient profile is the following: t_0, B= 5%; t_{15}, B= 65%; t_{16}, B=100%; t_{22}, B= 100%; t_{23}, B= 5%; t_{30}, B= 5%. Injected sample volumes were 5 µL.

In this MS feature the following parameters were set: MS and MS/MS scan speed 6 and 8 spectra/s respectively, for a total cycle of 1.26 s, 8 compounds per cycle, isolation width of 4 *m/z*, active mass exclusion enabled after 100 spectra and 0.3 min, absolute and relative precursor threshold 200 counts and 0.001 % respectively.

Simultaneous MS and tandem MS analysis were performed by using the AutoMS/MS feature. Centroid MSMS spectra were recorded over the range 50-2000 *m/z* with a scan rate of 3 spectra/s in both MS and tandem MS mode for a total cycle time of 1.26 s, 2 compounds per cycle, isolation width of 4 *m/z*, absolute and relative precursor threshold of 200 counts and 0.001 % respectively. Collision energies are reported in Table 9.

A ".csv" file including 82 PDE-5 inhibitor analogues was used as preferred list of precursor ions in the AutoMS/MS experiments. The corresponding Metlin database was generated and loaded within the Molecular Features Extraction (MFE), setting the following parameters and thresholds: a) ion compound filters ≥1000, b) retention time ≤ 2.5% of tolerance with respect to the seven selected PDE-5 Inhibitors standards and c) MME ≤ 20 ppm.

Mass spectra acquisition and data analysis was processed with Masshunter Workstation B 04.00 software (Agilent Technologies).

5.3.2.1 Quantification of PDE-5 and evaluation of the method performance.

Due to the lack of an internal standard, the quantification for target analytes was performed using the corresponding external calibration curve, using least squares regression. Precursor ions at the MS level were chosen as quantifier ions, after checking the simultaneously presence of at least one fragment ion with S/N >3 in the MS/MS acquisition mode.

The method performance was evaluated in terms of linearity, limit of detection (LOD) and quantification (LOQ), recovery, repeatability, and matrix effect following the guidelines reported in SANCO/10684/2009 [85] and Decision 2002/657/EC [86].

Five-point calibration curves were constructed by injecting standard solutions prepared in the extractant solution. Concentrations ranged between the limit of quantification to 1000 pg injected. LODs and quantification LOQs were experimentally estimated through a specific calibration with at least four levels at concentration close to the LOD, and applying a recent statistical method reported in literature [87, 88].

For herbal bulk analysis, trueness was calculated from replicate analysis (n = 3) of blank matrix spiked at three different concentrations levels 1.0, 5.0, 10.0 μg/g, comparing sample peak areas with those of standard solutions at the same concentration. Repeatability of the method was assessed as intra-day precision (n=3) expressed as relative standard deviation (RSD) of the peak areas, while reproducibility as inter-day precision (N=15).

Matrix effects were evaluated by comparing the slopes of three-point calibration curves obtained fortifying each sample before sample injection, with those of the standard calibration. Solvent and procedural blank samples were used to check selectivity of the method, and in particular, the absence of any carryover effect was verified by repeating injections of solvent blank after analysis of 1.0 ng of standard solution.

5.3.3 Direct infusion full scan protocol for molecular formulas identification in $PM_{2.5}$.

This protocol allowed the determination of the chemical composition of the sample through a fast data treatment.

Instrumental analyses were performed using high-resolution LTQ Orbitrap™ Velos mass spectrometer (Thermo Fisher, Bremen, Germany). The mass analyser was calibrated with the manufacturer calibration solution before each analysis. The mass accuracy of the instrument, checked before analysis, was always below 0.5 ppm. The instrument mass resolution was set at 100 000 (measure at *m/z* 400). Each sample was analysed in the ranges of *m/z* 100−650 and *m/z* 150-900, acquiring each range three times for 60seconds. The acquisition was considered acceptable only if the spray resulted sufficiently stable, with variations of the total ion current (TIC) profile versus time within 80- 100%.

NanoESI mass analyses were performed using a TriVersa Nanomate robotic nano-flow chip-based ESI (Advion Biosciences, Ithaca NY, USA) source. The direct infusion negative nanoESI parameters were as follows: ionization voltage 1.6 kV, back pressure

0.8 psi, capillary temperature 275 °C, S-lens RF level 60%, sample volume 8µL. For the positive mode the same parameter were used except ionization voltage and back pressure set at 1.4kV and 0.3 psi respectively.

APPI analyses were performed using an Ion Max™ source (Thermo Fisher, Bremen, Germany) set to work in APPI mode with a Syagen Krypton lamp emitting photons at 10.0 eV and 10.6 eV. Source parameters were: temperature 200C, auxiliary gas flow 5a.u. (arbitrary units) and sweep gas 10 a.u. The flow used in direct injection was 10µL/min.

The mass analyser was calibrated before the analysis on the samples using the commercial calibration solution. The mass accuracy of the instrument was checked before the analysis and was below 0.5 ppm.

5.3.3.1 Post run data analysis

The post-run data analysis workflow for the assignation of unique molecular formulas to each *m/z* values involves four steps here summarized:

1. Assignation of molecular formulas to the experimental MS signals.
2. Determination of noise, mass precision and mass accuracy of the measures.
3. Molecular formulas filtering.
4. Determination of the common ions in the replicate acquisitions.

The first step was carried out using the qualitative browser of the software Xcalibur™ 2.1 by Thermo Scientific, which can generate molecular formulas for the peaks present in the MS spectrum, enforcing the results to observe some constriction described below. The steps 2 and 3 were implemented through two separate algorithms ("Mass Drift v1.11" and "MassSpecProcessing v1.0" respectively) wrote in Mathematica programming language and implemented in Mathematica 10.0 by Wolfram Research. Steps 2 and 3 were separated in order to perform a manual control of the process and identify possible failure in the "Mass Drift v1.11" algorithm. Finally, using Excel, mass ranges are merged in the same spreadsheet and the common ions determined in the replicates are obtained. Each mass spectrum was obtained by the average of one minute of acquisition, corresponding at 40 single spectra. In the generation of molecular formulas, carried out in Xcalibur 2.1 qualitative software, the follow constrains on the elemental composition were applied: $1 \leq {}^{12}C \leq 75$; ${}^{13}C \leq 1$; $1 \leq {}^{1}H \leq 180$; $1 \leq {}^{16}O \leq 50$; ${}^{14}N \leq 30$; ${}^{32}S \leq 2$; ${}^{34}S \leq 1$, mass tolerance 6ppm and maximum number of formulas per peak 10. For positive nanoESI acquisitions, the presence of one sodium atom is allowed in the molecular formula generation. Limiting the number of generated possibility was

needed to avoid extremely long processing time. The formulas list associated to each peaks, are exported in CSV file containing the accurate *m/z* value (five decimal digits), the intensity of the signals, the associated possible molecular formulas and the mass errors referred to the exact mass of the formulas expressed in ppm.

Mass error offset (mass accuracy, μ) and standard deviation of mass errors (mass precision, SD) were automatically calculated using the algorithm "Mass Drift v1.11", on the basis of known contaminants or substances likely to be present in the sample and previously confirmed via MS/MS experiments.

The algorithm "MassSpecProcessing v1.0" filters the assignments on the bases of heuristic rules. The elemental ratios were set at H/Cmin 0.3; H/Cmax 2.5; O/C max 2; O/Cmin 0; N/Cmax 1.3; S/Cmax 0.8. Signal to noise level was set at 5.

The common ions in the replicates of the two ranges are then merged in the same excel spreadsheet and the duplicates entries were removed.

6 Development of a two-step protocol for target and suspect analysis of freshwater cyanotoxins by LC/Q-TOF system.

6.1 Protocol optimization

6.1.1 Creation of database.

When LC/HRMS is used for suspect screening and confirmation of target compounds with reference standards, the positive candidate has to comply with general queries, *i.e.* matching retention times and/or fragmentation pattern with acceptable tolerance, and a signal or S/N higher than a certain threshold[89]. The possibility of identify and/or quantify suspect contaminants is limited by the creation of databases as large as possible, with retention times and accurate MS and MS/MS spectra of target compounds. Because the non-availability of a fit-to-purpose commercial or online database, the development of an in-house library was required.

Molecular formulas of selected cyanotoxins with relative monoisotopic mass and chemical structures, uploaded as .mol files, were inserted in the library, for a total of 210 entries, belong to MCs (110), anabaenopeptins (ANPs) (23), cyanopeptolins (13), microginins (9), saxitoxins, anatoxins, cilindrospermopsins and other oligopeptides (the full list is not reported). Some MC-related compounds reported as transformation products by water treatment [90] were added to explore their contribute to drinking waters contamination. All entries were found in the literature and all the references are not reported because the great number. Moreover, the database could be improved in terms of number of entries; retention times and fragments emerged from cyanotoxins characterization.

6.1.2 Optimization of the instrumental conditions

An initial instrumental condition was adopted according a previous study [12]. Then, two different columns, both based on the "solid core technology", namely Kinetex C-18 and Accucore (2.6µm 100 mm x 2.1 mm i.d, Superchrom, Italy) were evaluated in terms

of peak width and resolution on selected standards of cyanotoxins, using ESI source in positive acquisition mode. Mobile phase composition was modified to improve peak shapes and MS signals intensity of the selected cyanotoxins. Water, acetonitrile and methanol all containing formic acid at different concentrations (10 mM and 0.1% v/v) were tested on both columns. Flow rate, initially set at 0.2 mL/min, was increased up to 0.3 mL/min in order to improve S/N and consequently the Limits of Detection (LODs) of the methods, without incurring in the signal depletion due to a poor efficiency of the ionization process. Results of these experiments have shown that the best conditions in terms of sensitivity, S/N and chromatographic efficiency were obtained with the Kinetex C-18 column, using water and acetonitrile; both contains formic acid at 10 mM, at flow rate of 0.3 mL/min.

For MC variants that exhibit both the mono charged molecular ion $[M+H]^+$ and the doubly charged molecular ion $[M+2H]^{2+}$, the most intense between the two signals was used in the subsequent processing. Figure 9 shows the extracted ion current profile for a working standard solution of available MCs injecting 250pg of each compounds. The only remarkable observation about ESI-MS pattern of the MCs analysed, is the presence of the double charge protonated molecular ion $[M+2H]^{2+}$ as base peak for the MC-LR (*m/z* 498.2817), [D-Asp^3] MC-LR (*m/z* 491.2738), and MC-YR (*m/z* 523.2713), while the $[M+H]^+$ ones account for about 40%. Conversely, MC-RR variants exhibit almost exclusively $[M+2H]^{2+}$and the other selected congeners only $[M+H]^+$ ones [91]. This general behaviour has been useful for the characterization of the non-target MCs by using MS fragmentation experiments.

Negative acquisition mode was evaluated using ammonium as modificant in the range 0.2-10 mM in both mobile phases, since the Accucore column can be used with a quite large range of pH (1.5-11). Although signals of acid variants of MCs, like MC-LA, MC-LY and MC-LW, have shown a significant improvement with the increasing concentration of ammonium, nevertheless this conditions were not adapt for analysing other cyanotoxins with basic amino acids (signals about one order of magnitude lower). Even using mixed mobile phases (acidic water and basic acetonitrile), and switching the acquisition mode from ESI+ to ESI- during the chromatographic run, results were not suitable to a comprehensive target method.

Identification and confirmation purposes need the presence for each structure present in the database of fragmentation spectra. As fist attempt, in-source fragmentation was studied. High transmission voltage (fragmentor voltage) have been used to produce the

in-source fragmentation in order to obtain exact masses of both precursor and daughter ions in a single chromatographic run. Experiments have been performed with some representative MCs, after the variation of the fragmentor voltage (range 100-350 V) to obtain both molecular ions and the characteristic ion fragment at 135.0804 *m/z*, related to the $[C_9H_{11}O]^+$ ion generated by the breakage of the common Adda amino acid. Data obtained have showed that this approach is not effective, since a unique value of fragmentor voltage was not suitable to produce molecular and daughter ions with the adequate sensitivity, for analysing all target and suspect cyanotoxins. In fact, different fragmentor voltages are required, due to the generation of mono and double charged molecular ions related to the amino acid composition, resulting in a different energy involved in the MS fragmentation process. Moreover, the unpredictable co-elution of MC variants could generate merged fragment ions related to the Adda-moiety that results in a non-specific assignment of the structure.

Due to the issues present in the use of in-source fragmentation, we have preferred to perform the fragmentation in the collision cell. The optimization of collision energies for the target MS/MS analysis, were made by using MC standards. Two ranges of collision energies were outlined: 15 eV for double charged and 60 eV for mono-charged molecular ions respectively. The general fragmentation behaviour of the MC congeners is the α-cleavage at the methoxy-group of the Adda β-amino acid moiety, with the production of the previously described ion at 135.0804 *m/z*, often together with its complementary $[M+H-134]^+$ [12].

Retention times, LODs, optimal collision energies, and product ions of the MC variants available are show in Table 6.

Table 6. List of selected MCs with corresponding molecular ion masses, retention times, collision energies and daughter qualification ions; *doubly charged species.

Compound	Theoretic molecular ion, *m/z*	Tr, min	LOD, pg inj	Collision Energy, eV	Theoretic fragment ion, *m/z*	Fragment structure
[D-Asp3]MC-RR	**512.7827*** 1024.5574	6.05	20	15	890.4724 135.0804	$[M+H-134]^+$ $[PhCH_2CH(OMe)]^+$
MC-RR	**519.7905*** 1038.5731	6.29	4	15	904.4981 135.0804	$[M+H-134]^+$ $[PhCH_2CH(OMe)]^+$
MC-YR	1045.536	7.9	50	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-HtyR	1059.548	7.99	50	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-LR	995.5564	8.11	30	60	213.0828 135.0804	$[Glu+Mdha]^+$ $[PhCH_2CH(OMe)]^+$
[D-Asp3]MC-LR	981.5402	8.16	50	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$$[PhCH_2CH(OMe)]^+$
MC-HilR	1009.574	8.47	50	60	213.0859 135.0804	$[Arg+NH_3CO]^+$ $[PhCH_2CH(OMe)]^+$
MC-LA	**505.2920*** 909.4848	8.55	30	15	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-WR	1068.553	8.71	40	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-LY	1002.517	11.62	30	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-LW	1025.536	12.98	40	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$
MC-LF	986.5226	13.21	30	60	375.2011 135.0804	$[C_{11}H_{13}O+Glu+Mdha+H]^+$ $[PhCH_2CH(OMe)]^+$

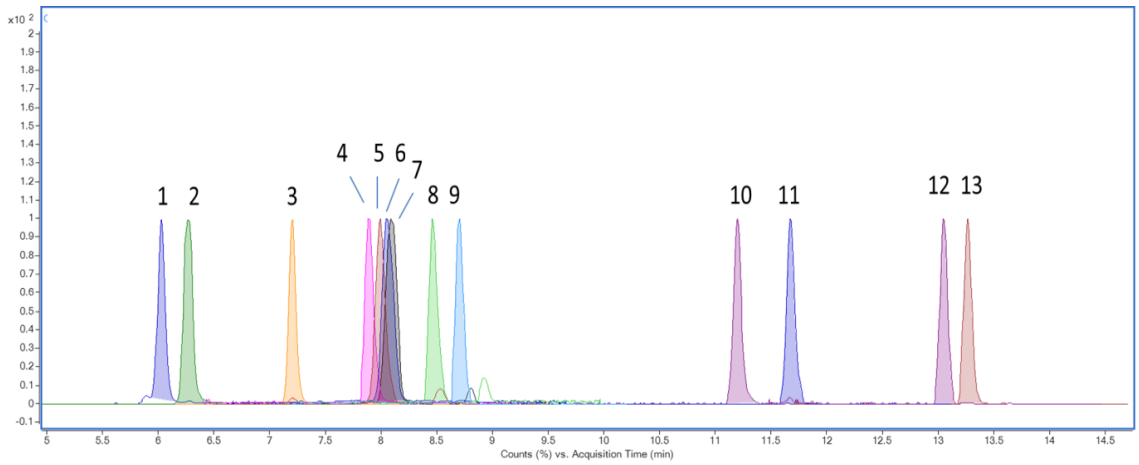

Figure 9. Chromatographic profile of 1µL of standard solution at concentration of 250pg/µL for MCs and 1ng/µL of NOD; 1) [D-Asp3]MC-RR 2) MC-RR 3) NOD 4) MC-YR 5) MC-HtyR 6) MC-LR 7) [D-Asp3]MC-LR 8) MC-HilR 9) MC-LA 10) MC-WR 11) MC-LY 12) MC-LW 13) MC-LF.

6.1.2.1 Post-run data analysis: full scan acquisition

A post-run data analysis approach allows the identification of suspect contaminants re-processing data acquired in full scan mode, without analyte-specific optimizations. This feature is suitable to be used, even in parallel with more robust quadrupoles, for simultaneously target confirming and suspect screening purpose. Consequently, data will be useful for selecting emerging contaminants, in the frame of food and environmental safety.

Thus, a database was created using the Metlin feature embedded with the Masshunter software that takes into consideration automatically multi-charged molecular ions.

Threshold values and queries were set within the MFE functions to narrow the list of positive candidates (see experimental section). Chromatographic signals with retention times lower than two minutes were rejected as a further query, since in this range the accuracy and consequently, the reliability of results, was strongly influenced by matrix compounds. It is worth noticing that the tolerance of 20 ppm is not so large, if we consider that 10 mDa value, often used for the analysis of low-weight pesticides, corresponds to about 15-150 ppm for that molecular weights. The value chosen by us, corresponding to about 20 mDa for the majority of cyanotoxins, includes all the sources of uncertainty related to non-target analysis in real samples, e.g. analyses at trace levels, saturation of the detector and matrix effects.

Nevertheless, the number of total positive results was huge, even because all water samples were selected among ones highly contaminated by cyanobacteria. Thus, about two hundreds of results were carefully checked to eliminate false positive results due to

a) ions recognized as salts adduct of the databases entries, but actually related to different signals; b) ions clearly referable to series of homologues compounds, e.g. differing for an ethoxylate group; c) ions belonging to the isotopic pattern of other signals; d) ions ascribable to other compounds loosing water or ammonia. Finally, the ratio between $[M+H]^+$ and $[M+2H]^{2+}$ was considered a remarkable criterion for the identification of toxins containing basic structures like Arg (R-series of MCs). Following this scheme, the initial list was substantially reduced to a total of about a hundred hits for all samples processed. Then, all results related to algal toxins for which standards are not available were subjected to a structural confirmation by means of tandem MS analysis in target acquisition mode.

6.1.2.2 Post-run data analysis: MS fragmentation

CID fragmentation was used to confirm the toxins which the standard is present and the standardless metabolites. Two ranges of collision energies were outlined, slightly different from those used for MCs optimization: 10-20 eV for $[M+2H]^{2+}$ and 40-70 eV for $[M+H]^+$ respectively. Only compounds confirmed by the presence of two or more characterizing fragment ions were considered positively identified (Table 8).

A general behaviour of the MCs congeners is the α-cleavage at the methoxy group of the Adda β-amino acidic moiety, with the production of the previous described ion at 135.0804 *m/z*, often together with its complementary $[M+H-134]^+$ [92]. Similarly, ANPs exhibit a typical signal at 263.1390 *m/z* , corresponding to the $|MAla^2\text{-}Hty^3|$ charged residue (through all the text dm stands for demethyl, *m* for O-methylation, while M for N-methylation). Figure 10 and Figure 11 report the general structures of MCs and anabanopeptins with their corresponding class-fragment ions.

Other significant signals were ascribed to the typical fragmentation pattern, mainly *b* and *y*-type ions, with relevant difference due cyclic nature of these peptides and the involving of the peculiar Adda residue [93][94]. Further information has been supplied in the specific discussion of algal metabolites encountered in water samples.

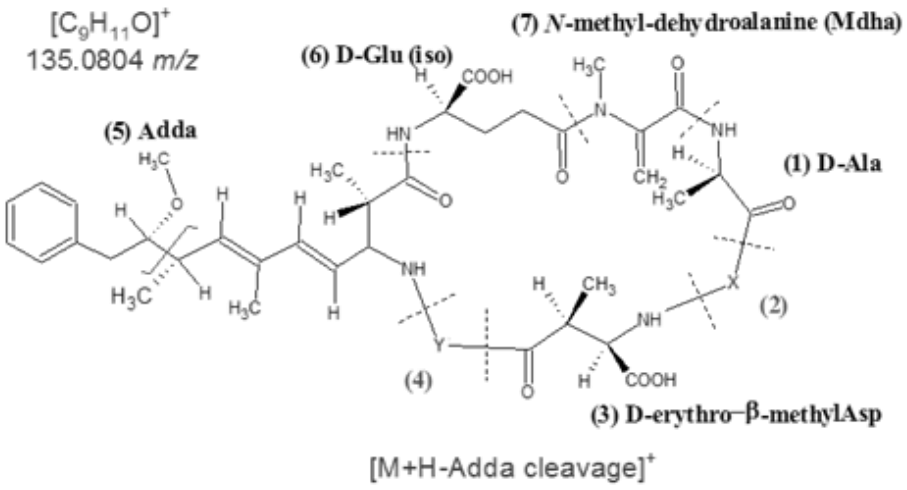

Figure 10. General structure of microcystins with related class-fragment ions.

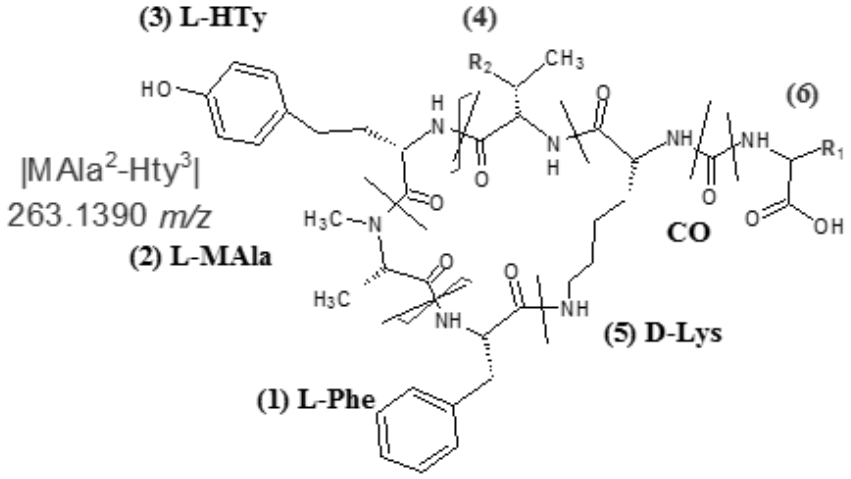

Figure 11. General structures of Anabaenopeptin with related class-fragment ions

6.1.3 Auto MS scan

An alternative approach for cyanotoxins identification is represented by the possibility to perform an automatic fragmentation of relevant signals acquired in HRMS full scan mode within a single chromatographic run. When the Metlin file was converted into a .csv file to be used as a preference list, attention was taken at adding both $[M+H]^+$ and $[M+2H]^{2+}$ according to the observed MS behaviour. The instrumental parameters described in the experimental section have been optimized using standard solutions and real samples. In particular, the scan rate in MS level and MS/MS level and number of ions processed in each cycle were selected in order to have a good peak reconstruction (needed for the quantitation) and the presence of the fragmentation spectra maintaining

at the same time a good sensitivity for all target and non-target compounds. Active exclusion was introduced to in the method because, in the first attempts, several peaks were lost.

Since the presence of several algal toxins in a limited range of retention times, requires the rapid acquisition of a high number of MS and MS/MS spectra. In this circumstance, the typical MCs fragment ion is not sufficient to guarantee a specific characterization, since co-eluted MCs variants with the same precursor ion (within the isolation width range), could result in merged MS/MS signals, not easy to be interpreted. The auto MS scan feature was evaluated in terms of number of positive results at the MS level, presence of diagnostic daughter ions, chromatographic peaks definition and detection capability. Results have been compared with the two-steps acquisition mode: auto MS profiles obtained, appeared qualitatively equivalent to those arisen from the two-steps analysis, ensuring the accurate determination of both molecular and fragment ions. Moreover, MS/MS fragmentation is often available from both mono and double charged precursor ions, giving useful information on peptide residues. Anyway, a lower sensitivity was experienced for the auto MS scan, being its MS signals 50-80% lowers than corresponding ones in MS mode. This fact influenced the number of positive results, when the supposed algal toxins were present at very low concentrations (low MS counts). Consequently, up to 10% of identification ability was lost for water samples contaminated at so critical concentrations. Comparable results were obtained by applying the internal database in post-run analysis of the autoMS/MS file. Notwithstanding the described drawbacks, auto MS approach seems to be more friendly in terms of time of analysis and data interpretation, because the presence of the typical class-fragments for MCs and ANPs is useful to identify these compounds, without any other consideration about possible interferences.

6.1.4 Quality control and quantitation

Quality control has been mainly focused on the instrumental accuracy. The use of a continuous infusion of the manufacturer calibrant solution is not suitable, because a significant matrix effect can affect identification and quantitation [95]. Thus, the use of a lock mass (diisooctyl phthalate) has been preferred. After an initial tune calibration, the MME of nodularin, always added as IS at 1 µg/L [12], was 5.67±3.22 ppm assessed on 21 measures over three working-days. This performance, without any drift observed,

suggested that this raw approach is suitable for confirmatory-screening purpose, if further criteria, like MS fragmentation and retention times for target analytes, were considered. Retention times variations for IS and cyanotoxins with standards were lower than 0.1 min for all analyses performed.

Calibration of selected cyanotoxins has been performed on MS levels to explore the linear dynamic range of the LC-MS system. Coefficients of variation R^2, obtained injecting in duplicate 20 µL of standard solutions in the 0.25-100 ppb range, were always close to the unit (Table 7), while overall RSD were lower than 5%. No matrix effects have been observed.

For the available standards, LODs has been estimated by evaluating S/N of the extracted ion current related to the exact molecular ion mass with a tolerance of 20 ppm from the full-scan spectra. LODs, assessed as the concentration equal to S/N=3 from injecting 50 pg of a standard solution, were reported in Table 7. Devoted validation experiments and a robust statistical analysis are necessary for any other specific application to be implemented, *i.e.* food analysis. These values were proposed only as an indication of the instrumental possibility of this approach.

Table 7. Calibration parameters and LODs of selected toxins.

	External calibration		**Internal Standard calibration**		**LOD (ppb)**
	regression equation	R2	regression equation	R2	
[D-Asp3]MC-RR	y = 19196x - 28768	0.9989	y = 0.0036x - 0.0043	0.9979	20
MC-RR	y = 9282x - 8454.5	0.9953	y = 0.0017x + 0.0004	0.9969	4
MC-YR	y = 949x - 475.63	0.9983	y = 0.0002x + 0.0002	0.9973	50
[D-Asp3]MC-LR	y = 1864x - 2017.8	0.999	y = 0.0003x + 0.0001	0.9986	50
MC-LR	y = 1474x - 1511.7	0.9989	y = 0.0003x + 0.0002	0.9982	30
MC-LA	y = 2058x - 330.17	0.9988	y = 0.0004x + 0.0004	0.9995	30
MC-LY	y = 1897x - 1225	0.9985	y = 0.0003x + 0.0003	0.9996	30
MC-LW	y = 1655x - 4031.4	0.998	y = 0.0003x - 0.0003	0.9986	40
MC-LF	y = 2107x - 2934.7	0.9985	y = 0.0004x – 0.0001	0.9994	30

6.2 Algal toxins identification in freshwaters

We have to point out that results on suspect identification of cyanotoxins here described are preliminary, since the analyses of surface and drinking waters were performed on a limited samples volume. Quantitation of the cyanotoxins with standards, already performed by ISS, was out of our scope, and the MS characterization of the suspect cyanotoxins has been restricted to the most relevant compounds, emerged from database

or auto MS scan processing. Anyway, the sample named Bidighinzu, collected in a lake of t Sardinia isle, was hardly contaminated by *Microcystis* spp and *Aphanizomenon* spp (about 10^{11} cells/L, reported by official analysis). The direct analysis of an aliquot of 40 µL of the filtered water sample, allowed to achieve very high MS signals (often $>10^5$) of compounds subsequently recognized as cyanotoxins. Thus, multiple MS/MS analyses for characterizing algal toxins were conducted mainly on this sample.

6.2.1 Microcystis toxins congeners

All water samples collected were affected by cyanobacteria, belonging to the *Microcystis aeruginosa, Planktothrix rubescens* and *Aphanizomenon flos aquae*. A description of water samples with results obtained with LC/HRMS analysis is reported in Table S 1, whilst Table 8 shows a detailed elucidation of the ions significant for the structural elucidation of selected target algal toxins.

Twelve uncommon MC congeners were identified using both LC/HRMS protocols proposed, derived from MFE processing followed by target MS experiments, or from auto MS scan. We have here focused our attention on product ions characterizing the specific MC variant.

Three high signals at 500, 507 and 532 *m/z*, ascribing at $[M+2H]^{2+}$ species by the isotope patterns, and with correlated MS/MS ions, retained between RR and LR series, were recognized as two $MSer^7$ variants of the MC-LR and one of the MC-YR [96]. Fragment ions accountable to MC-LR having the amino acid in position 3 methylated (*m/z* 399.2350, 416.2616 and 841.4818) or demethylated (272.1353, 385.2194, 402.2459 and 637.3596) were observed. The lack of the major signals involving $|Mdha^7\text{-}Ala^1|$ residues (e.g. 155.0815, 397.2082), always observed by us with MC standard, confirmed a modification in positions 1-7. The assignment to $MSer^7$-LC variants was supported from several fragments experienced by us, *i.e.* those at *m/z* 173.0921, 771.4400, attributed to the sequence $|MSer^7\text{-}Ala^1|$ and $|Arg^4\text{-}Adda^5\text{-}Glu^6\text{-}MSer^7\text{-}Ala^1|$ respectively. Anyway, we found signals elsewhere attributed to the conventional sequence $Mdha^7\text{-}Ala^1$ (*m/z* 213.0870 and 375.1915), in spite of those considered diagnostic for the $MSer^7$ residue, *i.e.* ions 393 and 231 [93], which were not observed by us, neither using high nor in low collision energies. These fragments could be ascribed, even employing the useful freeware software *mmass*, to a dehydration of the corresponding *b*-type ion.

The same fragmentation pattern was observed for a compound eluted early before the described [MSer^7] MC-LR variants, which we have characterized as the [MSer^7] MC-YR. We have to point out that its molecular ion could be hypothetically referable to the [Ser^7] MC-HtyR, as indicated by records of our database, but the MS behaviour was not coherent with this structure. This fact often occurred with putative MC signals.

Two signals corresponding to a demethylation of the MC-YR was recorded at t_r=6.33 and 7.80 minutes. For the former peak, the presence of the fragment at *m/z* 121.0648 instead of the conventional 135.0804, and other signals indicating the methylation in position 7 and 3 (ions 213.0870, 375.1899 and 269.1244 respectively, data not shown), suggests a dmAdda^5 variant. The latter was identified as the [Asp^3] MC-YR, mainly for the presence of previously mentioned residues related to the sequence |Glu^6-Mdha^7-Ala^1| and to |Arg4-Asp3| (measured *m/z* 272.1364, Δ = +1.1 mDa). Both compounds have produced Tyr immonium ions [97].

Again, three isobaric congeners corresponding to a demethylation of the MC-LR, although with very low signals, result to be efficiently separated (Figure 12).

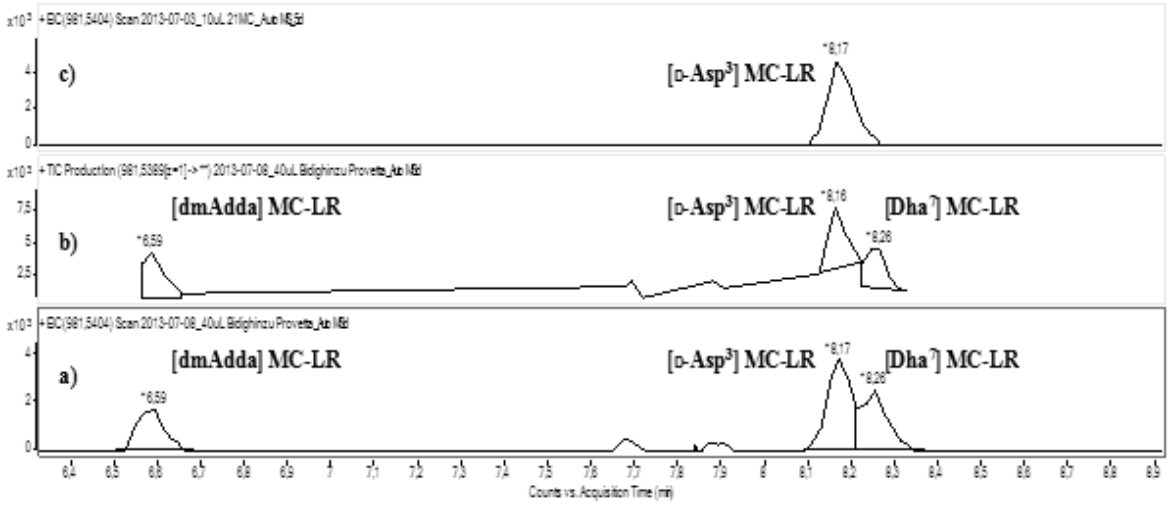

Figure 12. Chromatographic profiles of isobaric variants of demethyl (dm) MC-LR occurred in a water sample; a) MS Extracted Ion Chromatogram (EIC) of the theoretical [M+H]+; b) tandem MS TIC (Total Ion Current) of the [M+H]+ precursor ion; c) MS EIC of 250 pg injected.

In this case, the comparison with the retention time related to the standard of the [D-Asp^3] MC-LR, help us to assign this variant to the peak with higher intensity at t_r=8.17. The presence or absence of the diagnostic fragments ions [93] related to the Adda-cleavage (*m/z* 135.0804 and 847.4672 or 121.0648), Mdha^7 (*m/z* 127.0866, 213.0870 and 375.1899), Asp^3 (*m/z* 272.1353) or Dha^7 (*m/z* 475.2187), allowed us to characterize

these variants. Thus, we have assigned *via* MS fragmentation the structure of [dmAdda] MC-LR and [Dha^{7}] MC-LR to peaks at t_r=6.59 and 8.26 respectively, coherently with the chromatographic behaviour.

MC-M(O)R, with theoretical mass of 1028.5001 Da corresponding to $C_{48}H_{72}N_{10}O_{13}S$, was identified at t_r =7.26, from both $[M+H]^+$ and $[M+2H]^{2+}$ with of MME <3 mDa. The daughter ions at *m/z* 965.5091 and 803.4120, obtained from fragmentation of the $[M+H]^+$ at 60 eV, were ascribed to a characteristic loss from the methionine-S-oxide^{2} (MetO^{2}) and amino acidic residue of |Glu^{6}-Mdha^{7}| respectively. Another MS signal at *m/z* 682.3803 could be assigned to the sequence |Arg^{4}-Adda^{5}-Glu^{6}-Mdha^{7}| [98]. No signals referred to demethylation in position 3 or 7 were registered. These MS signals, together with those generally produced by the MC class, allow the characterization of this MC congener.

Although present with low counts, the structure of MC-FR has been assigned to a peak present in the Bidighinzu sample at t_r=8.72. The $[M+H]^+$ measured at 1029.5400 *m/z* (-0.4 mDa) was significantly different from that described for the MC-M(O)R, and it is potentially referable to other two variants of MCs, *i.e.* [Asp^{3}] MC-HphR and [Dha^{7}] MC-HphR. The Adda-related MS fragment at *m/z* 375.1899 showed the presence of amino acidic residues of |Glu^{6}-Mdha^{7}|. Besides, signals corresponding to Arg^{4} and Phe-immonium ion with theoretical 120.0808 *m/z* [99] have allowed us to assign this compound to the MC-FR variant in place of the isobaric demethylated forms of MC-HphR.

Other relevant signals that showed several MC typical daughter ions with good accuracy suggest the presence of several MC variants reported in literature [100]. Anyway, the fragmentation pattern was quite different from that expected, so that a deeper investigation is necessary in order to identify these compounds.

No signals corresponding to cyanotoxins by-products reported elsewhere [90] were evidenced, although several samples of contaminated drinking water treated with hypochlorite or chloride dioxide have been examined.

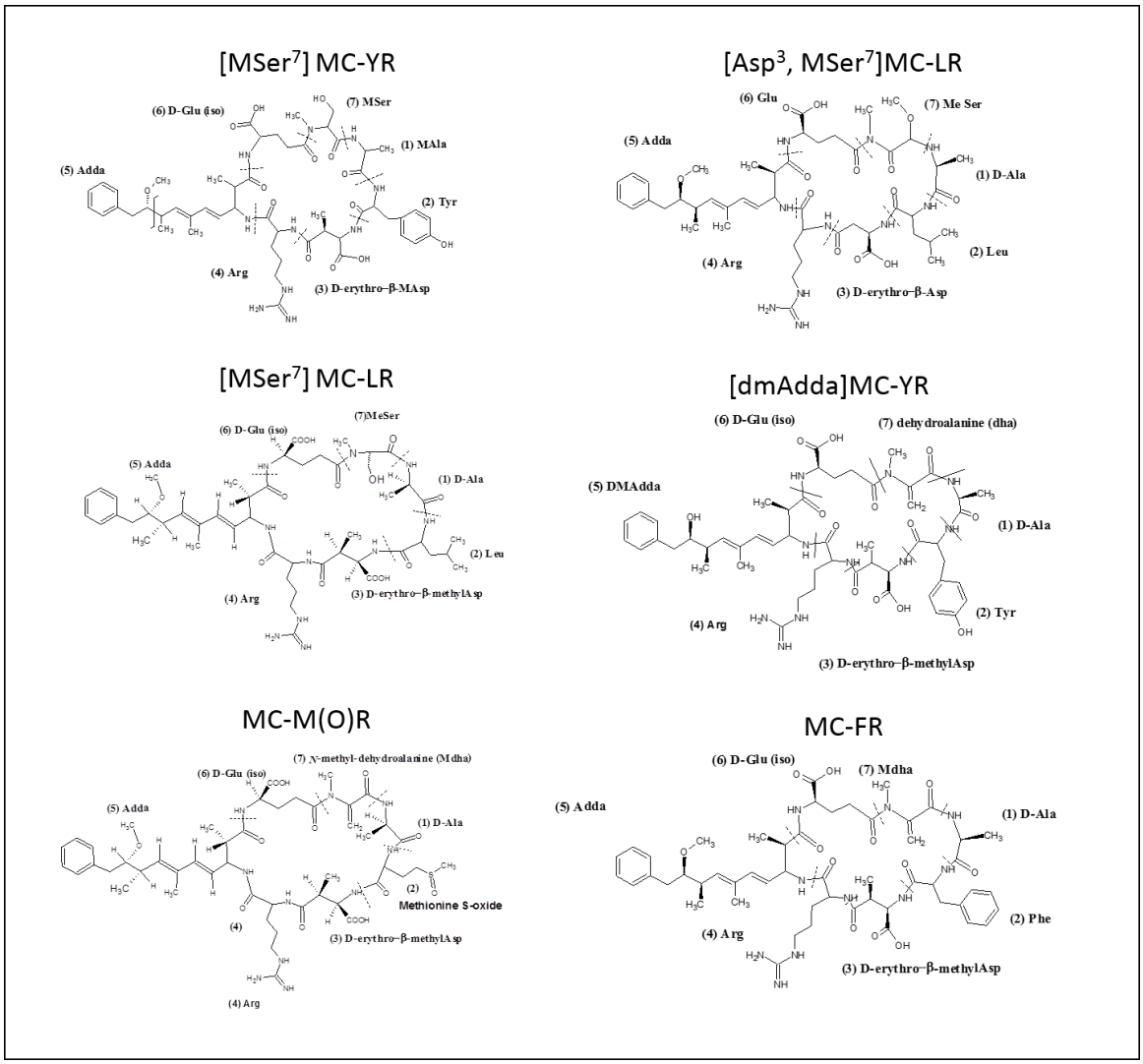

Figure 13. Molecular structures of identified uncommon Microcystis.

6.2.2 Anabaenopeptins.

About 30 different ANPs, having a general hexapeptidic structure of five cyclic amino acids, with main variations in residues 6 and 4, and a characteristic uredo linkage have been described (Figure 10). Their toxicity, at concentrations occurring in environment, is currently unknown, although their biological activity includes inhibition of PP and serine protease [93, 101, 102]. ANP–type peptides have been produced from several cyanobacterial species, including *Microcystis* spp, *Planktothrix* and *Aphanizomenon* spp [93] occurred in water samples analysed by us. Data are reported in Table S 1 and Table 6.

ANP-A and B were found in several water samples at t_r =7.6 and t_r =4.7 minutes respectively, as both $[M+H]^+$ and $[M+2H]^{2+}$, with different relative intensities. In fact,

ANP-B, having an Arg instead of a Hty in position 6, gives a $[M+2H]^{2+}$ as base peak. Fragmentation of the two substances, performed on the $[M+H]^{+}$ precursor ion at 40 eV for ANP-A and on $[M+2H]^{2+}$ at 10 eV for ANP-B, gave the class-fragment at *m/z* 263.1390 and a number of product ions already described in literature [101, 103]. Their presence in water samples could be confirmed by comparing retention times and MS data with those of respective standards, which are nowadays commercially available, but did not for us at the time of this research.

Three isobaric ANPs variants (F, E, B1 or MM850) with molecular formula $C_{42}H_{62}N_{10}O_9$ and a molecular weight of 850.4701 Da, that cannot be distinguished with a MALDI-TOF system in full scan analysis [93], have been reported [98, 104, 105] (Figure 15). We often found a signal at t_r =5.5 min related to their $[M+H]^{+}$, that we assigned to ANP-F *via* fragmentation pattern. Instead, two distinct signals resolved with high efficiency (t_r =5.51 and t_r =5.99 in Figure 3), were observed in surface samples contaminated by *Planktothrix rubescens* and *Microcystis aeruginosa* or *Plankthotrix rubescens* and *Aphanizomenon flos-aquae* (Table S 1). Both compounds exhibit $[M+H]^{+}$ and $[M+2H]^{2+}$. Subsequently, in the auto MS scan mode, it was possible to obtain the fragmentation pattern of the two isobaric compounds in one step, using $[M+H]^{+}$ (CE 40 eV) and $[M+2H]^{2+}$ (CE 10 eV) as precursor ions. The assignment of the second signal to ANP-E having a MeHty in position 3 [101] was excluded, since fragments related to the Hty immonium ion and its amino acidic loss (theoretical *m/z* 674.3984) were always present. Moreover, this compound showed MS/MS pattern similar to ANP-B, having Val^4 instead of Ile^4 (theoretical *m/z* 637.3708), but with an increment corresponding to a methylation for all fragments containing the Arg residue (Figure 8 and Figure 14). The existence of an ANP named B1, presumably containing Har (homoarginine) in position 6, was supposed by Ferranti *et al.* [104], but a structure coherent with such fragmentation pattern was previously described for the ANP-MM850. The structure of this ANP-MM850, exhibiting a methyl ester of the Arg^6, was confirmed by NMR analysis. Thus, we are prone to sustain the last description for the ANP variant experienced by us. Coupling the chromatographic separation to a "double" fragmentation process consented the unambiguous discrimination of these isobaric ANPs.

Another ANP, with fragment ions referable to the presence of Ile^4 (*m/z* 651.3794) and Har^6/*m*Arg^6 (*m/z* 189.1337 and 215.1131) was characterized at MS and MS/MS level. A similar structure with Har^6 has been previously named F1. Anyway, even in this case,

the structure could be ascribed to a modification of the ANP-F, with a *m*Arg[6]. Until the molecular structure will be confirmed with NMR data, this algal toxin will be here named ANP-MM864.

Oscillamide Y, a well known ANP-analogue, with formula $C_{45}H_{59}N_7O_{10}$ was present in several samples. In this case, structural information is available from fragmentation of the $[M+H]^+$ with a very low MME.

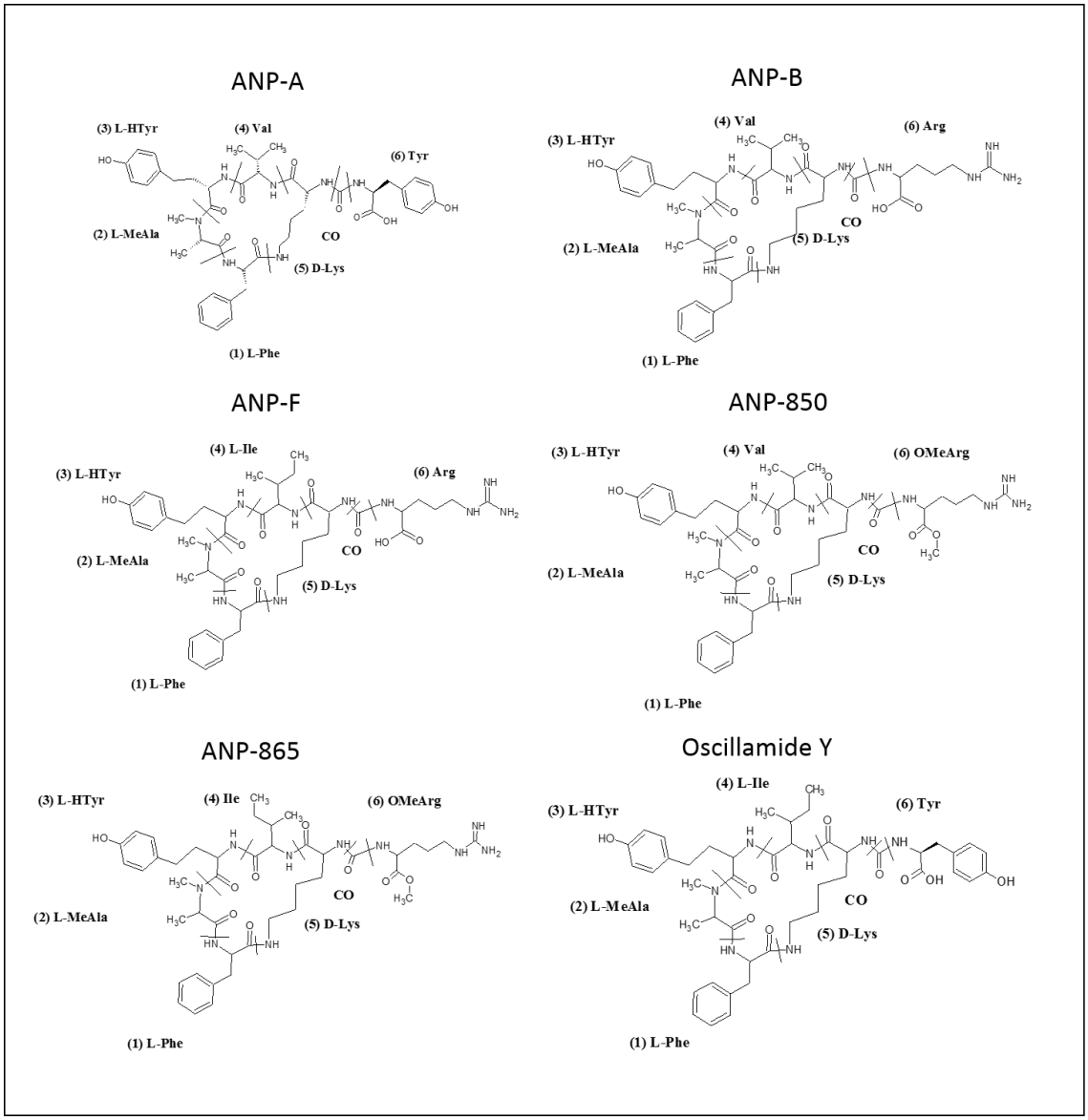

Figure 14. Molecular structures of the uncommon Anabaenopeptins identified.

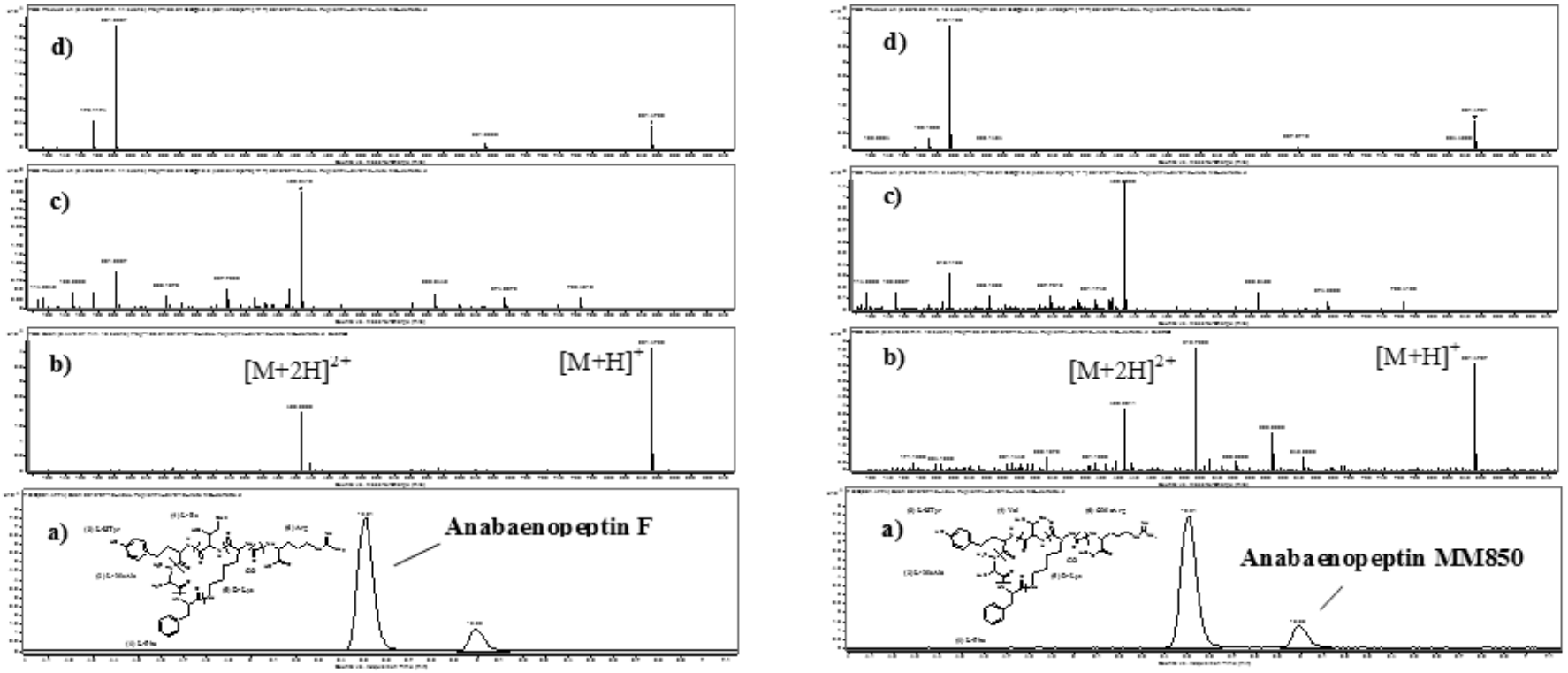

Figure 15. Extracted Ion Chromatogram (EIC) and relative mass spectra assigned to the isobaric anabaenopeptin F (left panel) and MM 850 (right panel), obtained from the analysis of a surface water sample from the lake Occhito: a) EIC at the MS level of the ion at at *m/z* 851.4774 with a tolerance of 20 mDa; b) mass spectrum at the MS level showing $[M+H]^+$ and $[M+2H]^{2+}$ ions; c) product ion scan obtained by using autoMS/MS features and $[M+H]^+$ as precursor ion; d) product ion scan obtained by using autoMS/MS features and $[M+2H]^{2+}$ as precursor ion.

Table 8. List of some representative non-target cyanotoxins confirmed by means of tandem MS analysis, with relative retention times, molecular and significant fragment ions, supposed structures and MME.

Compound	Tr, min	Molecular formula	Molecular ion, *m/z* (MME, mDa)[1]	Fragment ions, *m/z* MME, mDa)	Fragment formula[2]	Assigned structures
[MSer^7] MC-YR	6.18	$C_{52}H_{74}N_{10}O_{14}$	**532.2771**[3] (+0.5)	135.0798(-0.6)	$[C_9H_{11}O]^+$	$[PhCH_2CH(OMe)]^+$
				929.4687 (-4.0)	$[C_{43}H_{65}N_{10}O_{13}]^+$	$[M+H-134^4]^+$
				173.0911[6] (-1.0)	$[C_7H_{13}N_2O_3]^+$	$\|MSer^7\text{-}Ala^1\|$
			1063.5439 (-2.0)	891.4570 (-4.1)	$[C_{45}H_{63}N_8O_{11}]^+$	$\|Tyr^2\text{-}MAsp^3\text{-}Arg^4\text{-}Adda^5\text{-}Glu^6\|$
				771.4382 (-1.8)	$[C_{38}H_{59}N_8O_9]^+$	$\|Arg^4\text{-}Adda^5\text{-}Glu^6\text{-}MSer^7\text{-}Ala^1\|$
				466.2389 (-2.0)	$[C_{20}H_{32}N_7O_6]^+$	$\|Arg^4\text{-}MAsp^3\text{-}Tyr^2\|$
[Asp^3,MSer^7]MC-LR	6.41	$C_{48}H_{74}N_{10}O_{13}$	**500.2812 (+2.1)**	135.0802 (-0.2)	$[C_9H_{11}O]^+$	$[PhCH_2CH(OMe)]^+$
				865.4775 (-0.3)	$[C_{39}H_{65}N_{10}O_{12}]^+$	$[M+H-134]^+$
				402.2440 (-1.9)	$[C_{16}H_{32}N_7O_5]^+$	$\|Leu^2\text{-}Asp^3\text{-}Arg^4\|$
				173.0922 (+0.1)	$[C_7H_{13}N_2O_3]^+$	$\|MSer^7\text{-}Ala^1\|$
			999.5511 (+0.1)	771.4347 (-5.3)	$[C_{38}H_{59}N_8O_9]^+$	$\|Arg^4\text{-}Adda^5\text{-}Glu^6\text{-}MSer^7\text{-}Ala^1\|$
				385.2127 (-6.7)	$[C_{16}H_{29}N_6O_5]^+$	$\|Leu^2\text{-}Asp^3\text{-}Arg^4\|$
[MSer^7] MC-LR	6.51	$C_{49}H_{76}N_{10}O_{13}$	1013,5626 (-4.0)	135.0810 (+0.6)	$[C_9H_{11}O]^+$	$[PhCH_2CH(OMe)]^+$
				771.4408 (+0.8)	$[C_{38}H_{59}N_8O_9]^+$	$\|Arg^4\text{-}Adda^5\text{-}Glu^6\text{-}MSer^7\text{-}Ala^1\|$
				399.2381[6] (+3.1)	$[C_{17}H_{31}N_6O_5]^+$	$\|Leu^2\text{-}MAsp^3\text{-}Arg^4\|$
				375.1953[6] (+3.9)	$[C_{20}H_{27}N_2O_5]^+$	$\|C_{11}H_{13}O\text{-}Glu^6\text{-}MSer^7(\text{-}H_2O)\|$
			507.2869 (0)	879.5005 (+7.1)	$[C_{40}H_{67}N_{10}O_{12}]^+$	$[M+H-134]^+$

				173.0927[6] (+0.6)	$[C_7H_{13}N_2O_3]^+$	|MSer[7]-Ala[1]|
[dmAdda]MC-YR	6.33	$C_{51}H_{70}N_{10}O_{13}$	1031.5139 (-4.8)	269.1186 (-5.8)	$[C_{11}H_{16}N_4O_4]^+$	|MAsp[3]-Arg[4]|
				902.4784 (+1.3)	$[C_{46}H_{63}N_9O_{10}]^+$	|Arg[4]-dmAdda[5]-Glu[6]-Mdha[7]-Ala[1]-Tyr[2]|
				537.2762 (-1.8)	$[C_{23}H_{37}N_8O_7]^+$	|Ala[1]-Tyr[2]-MAsp[3]-Arg[4]|
			516.2692 (-5.2)	**456.2296** (-5.1)	$[C_{43}H_{64}N_{10}O_{12}]^{2+}$	$[M+2H-120^4]^{2+}$
				136.0740 (-1.7)	$[C_8H_{10}NO]^+$	$[im\text{Tyr}^5]^+$
				121.0612 (-3.6)	$[C_8H_9O]^+$	$[PhCH_2CH(O)+H]^+$
MC-M(O)R	7.26	$C_{48}H_{72}N_{10}O_{13}S$	1029.5053 (-2.1)	965.5102 (+1.1)	$[C_{47}H_{69}N_{10}O_{12}]^+$	$[M+H-CH_4SO]^+$
				803.4136 (+1.6)	$[C_{38}H_{59}N_8O_9S]^+$	|Ala[1]-MetO[2]-MAsp[3]-Arg[4]-Adda[5]|
			515.2597 (+0.4)	895.4334 (-0.8)	$[C_{39}H_{63}N_{10}O_{12}S]^+$	$[M+H-134]^+$
				135.0800 (-0.4)	$[C_9H_{11}O]^+$	$[PhCH_2CH(OMe)]^+$
MC-FR	8.72	$C_{52}H_{72}N_{10}O_{12}$	1029.5400 (-0.4)	895.4688 (+1.6)	$[C_{43}H_{63}N_{10}O_{11}]^+$	$[M+H-134]^+$
				135.0797 (-0.7)	$[C_9H_{11}O]^+$	$[PhCH_2CH(OMe)]^+$
			515.2766 (+2.8)	375.1899 (-1.5)	$[C_{20}H_{27}N_2O_{15}]^+$	|$C_{11}H_{13}O$-Glu[6]-Mdha[7]|
				174.1326 (-2.3)	$[C_6H_{16}N_5O]^+$	$[Arg^4+NH_3+H]^+$
				120.0785 (-2.3)	$[C_8H_{10}N]^+$	$[im\text{Phe}]^+$
ANP-A	7.56	$C_{44}H_{57}N_7O_{10}$	844.4232 (-0.8)	667.3396 (-5.4)	$[C_{34}H_{47}N_6O_8]^+$	|Val[4]-Lys[5]-CO-Tyr[6]-Phe[1]-MAla[2]|
				362.2036 (-3.8)	$[C_{19}H_{28}N_3O_4]^+$	|MAla[2]-Hty[3]-Val[4]|
				263.1371 (-1.9)	$[C_{14}H_{19}N_2O_3]^+$	|MAla[2]-Hty[3]|
ANP-B	4.73	$C_{41}H_{60}N_{10}O_9$	837.4616 (-0.2)	201.0970 (-1.2)	$[C_7H_{13}N_4O_3]^+$	|CO-Arg|
				637.3685 (-2.3)	$[C_{34}H_{49}N_6O_6]^+$	|Phe[1]-MAla[2]-Hty[3]-Val[4]-Lys[5]|
			419.2342 (-0.3)	263.1367 (-2.3)	$[C_{14}H_{19}N_2O_3]^+$	|MAla[2]-Hty[3]|

				752.4069 (-2.1)	$[C_{37}H_{54}N_9O_8]^+$	|Hty^3-Val^4-Lys^5-CO-Arg^6-Phe^1|
ANP-F	5.51	$C_{42}H_{62}N_{10}O_9$	**426.2399** (-2.4)	674.3970 (-1.4)	$[C_{32}H_{52}N_9O_7]^+$	|Ile^4-Lys^5-CO-Arg^6-Phe^1-MAla^2|
			851.4793 (+1.9)	201.0967(-1.5)	$[C_7H_{13}N_4O_3]^+$	|CO-Arg^6|
				651.3835(-3.0)	$[C_{35}H_{51}N_6O_6]^+$	|Phe^1-MAla^2-Hty^3-Ile^4-Lys^5|
ANP-MM850	5.99	$C_{42}H_{62}N_{10}O_9$	851.4757(-1.7)	637.3715 (+0.7)	$[C_{34}H_{49}N_6O_6]^+$	|Phe^1-MAla^2-Hty^3-Val^4-Lys^5|
				215.1125 (-1.4)	$[C_8H_{15}N_4O_3]^+$	|CO-*m*Arg|
				189.1383(-3.7)	$[C_7H_{17}N_4O_2]^+$	|*m*Arg|
			426.2311 (-11.2)	674.3920 (-6.4)	$[C_{32}H_{52}N_9O_7]^+$	|Val^4-Lys^5-CO-*m*Arg^6-Phe^1-MAla^2|
ANP-MM864	6.54	$C_{43}H_{64}N_{10}O_9$	865.4891 (-4.0)	651.3794 (-7.1)	$[C_{35}H_{51}N_6O_6]^+$	|Phe^1-MAla^2-Hty^3-Ile^4-Lys^5|
				215.1131 (-0.8)	$[C_8H_{15}N_4O_3]^+$	|CO-XArg^6|
				189.1337 (-0.9)	$[C_7H_{17}N_4O_2]^+$	|XArg|
			433.2469 (-5.3)	603.3596 (-1.7)	$[C_{29}H_{47}N_8O_6]^+$	|Ile^4-Lys^5-CO-XArg^6-Phe^1|
				688.4079 (-6.2)	$[C_{33}H_{54}N_9O_7]^+$	|Ile^4-Lys^5-CO-XArg^6-Phe^1-MAla^2|
Oscillamide Y	8.18	$C_{45}H_{59}N_7O_{10}$	858.4377 (-1.9)	681.3577 (-2.9)	$[C_{35}H_{49}N_6O_8]^+$	|Ile^4-Lys^5-CO-Tyr^6-Phe^1-MAla^2|
				376.2229 (-0.2)	$[C_{20}H_{30}N_3O_4]^+$	|MAla^2-Hty^3-Ile^4|
				263.1396 (+0.6)	$[C_{14}H_{19}N_2O_3]^+$	|MAla^2-Hty^3|

6.3 Conclusion

The ability of the proposed protocols to work both as a structural-based screening and as confirmatory method has been proved on surface and drinking waters. In all samples processed, the target determination of cyanotoxins with certified standards was successfully accomplished. The database specifically implemented demonstrated its potentiality for the identification and characterization of suspect algal metabolites. Auto MS scan and the two-steps post-run data analysis could be efficiently used, depending on the different planning and purpose of the water monitoring: the former, as one-shot analysis specifically devoted to elucidation of the cyanotoxins contamination; the latter, as a possibility to acquire general LC/HRMS data that can be processed *a posteriori* to find emerging contaminants, even different from algal toxins. If carefully optimized, the auto MS feature seems to be able in achieving both objectives, with an acceptable loss of sensitivity. Both approaches are suitable to be improved enlarging the database entries, even with characteristic fragment ions, and used at ultra-trace levels in whatever matrix involved with cyanobacteria (environmental, biological or food samples). For such purpose, a full validation will be necessary for quantitative analysis. Preliminary results have evidenced a complex scenario of contamination for freshwaters affected by cyanobacterial blooms, enriched by unrecognized potential toxic MCs, and oligopeptides with a doubt harmful bioactivity. Features of an LC/Q-TOF-MS coupled to information automatically gained by databases or libraries, can effectively improve knowledge about cyanobacteria metabolites occurring, and thus restyle the analytical tools available for risk management related to cyanotoxins exposure.

7 Development of “one-shot” analysis of PDE-5 inhibitors and analogues in natural products for the treatment of erectile dysfunction.

7.1 Optimization of the instrumental conditions.

The optimization of MS parameters for the seven PDE-5 inhibitors (fragmentor voltage, source parameters and collision energies) was performed in flow injection analysis (FIA) using individual standard solutions at 100 ng/mL at flow rate of 0.3 mL/min water:acetonitrile 20:80 with 0.1% FA

Different ion sources were tested both in positive and negative operating mode. The ESI probe operating in positive mode was preferred to APPI (Atmospheric Pressure Photo-Ionization) and APCI (Atmospheric Pressure Chemical Ionization) sources, owing to its better performances in terms of signals to noise ratio and its robustness.

The full scan acquisition of each analyte was made by varying the fragmentor voltage in order to 1) establish the most abundant ion, 2) obtain the best response and 3) investigate if an in-source fragmentation is possible. Quasimolecular ions $[M+H]^+$ were found to be the most abundant ones, being sodium and potassium adducts virtually absent. In two cases (Sildenafil and Pseudovardenafil) double charged molecular ions $[M+H]^{2+}$ were present, but $[M+H]^+$ were always chosen as precursor ions. No fragmentation was experienced in the range of fragmentor voltages tested, and a series of MS/MS experiments were implemented by varying the collision energy (CE) by 5 units (Figure 16) with the aim to ensure the largest response for both precursor and qualifier ions. The best collision energy was selected by the intersection of the fitting line of the quasimolecular ion to the one belonging to the qualifier ion. MS and MS/MS scan rate was also optimized accordingly in order to have at the same time the best response and a sufficient peak reconstruction. Table 9 shows the optimized fragmentor voltages, collision energies and related precursor and qualifier ions in MS/MS acquisition for the target compounds.

Table 9. Retention times, theoretical *m/z* for molecular and qualifier ions and optimized collision energies.

Compound	Molecular formula	RT	Molecular ion	Qualifier ion	Optimized CE
		(min)	(*m/z*)	(*m/z*)	(eV)
Yohimbine	$C_{21}N_{26}N_2O_3$	9.9	355.2016	144.0817	24
Tadalafil	$C_{22}N_{19}N_3O_4$	15.5	390.1448	268.1012	3
Pseudovardenafil	$C_{22}N_{29}N_5O_4S$	16.6	460.2013	151.0826	36
Sildenafil	$C_{22}N_{30}N_6O_4S$	11.4	475.2122	58.0666	32
Vardenafil	$C_{23}N_{32}N_6O_4S$	10.5	489.2279	151.0874	37
Homosildenafil	$C_{23}N_{32}N_6O_4S$	11.7	489.2279	72.0813	34
Hydroxyhomosildenafil	$C_{23}N_{32}N_6O_5S$	11.2	505.2228	99.0924	30

7.1.1 Optimization of the chromatographic conditions.

The optimization of chromatographic conditions was initially performed using a C18 column (Kinetex, 150mm × 2.0 mm I.D.; 2.6 µm particle size from Phenomenex, USA), thermostated at 30°C and using water and acetonitrile acidified eluent (25 mM FA) and a linear gradient from 10% of acetonitrile to 100% in 24 min. Signals of the composite working standard solution at 100 ng/mL were registered in full scan mode.

Heavy and persistent tailing end carryover phenomena limited to Vardenafil were always experimented, evidencing the presence of specific interactions between this analyte and the stationary phase, probably arising from the presence of free silanolic groups or metal impurities on the surface of the C18-phase. The extent of carryover phenomena was estimated to be 3 ng/mL equivalent in this condition.

In order to evaluate and reduce these interactions, the concentration of formic acid was varied in both water and acetonitrile from 10 mM to 40 mM, without success. Thus, trifluoroacetic acid (TFA) 0.1% *v/v* and pentafluoroproprionic acid (PFPA) 5mM was employed as ion pair agents in mobile phases. Anyway, while the peak tailing was eliminated, both modifiers did not reduce the extent of the carryover effect. Moreover, ESI response significantly decreased of about 15-40% for all analytes.

The active interactions between Vardenafil and the stationary phase were completely overcome by the use of a polymeric column PolimerX™ (Figure 17). The absence of silanols ensured no carryover phenomena using both TFA and FA as acidifiers in mobile phases. Formic acid was however chosen because it produces a greater response in ESI. In this optimized chromatographic conditions, the highest sensitivity and selectivity, in terms of chromatographic separation of the isobaric Homosildenafil and

Vardenafil, were reached. Figure 18 shows the chromatographic profiles and relative MS/MS spectra for target analytes at concentrations of 10 times LODs estimated.

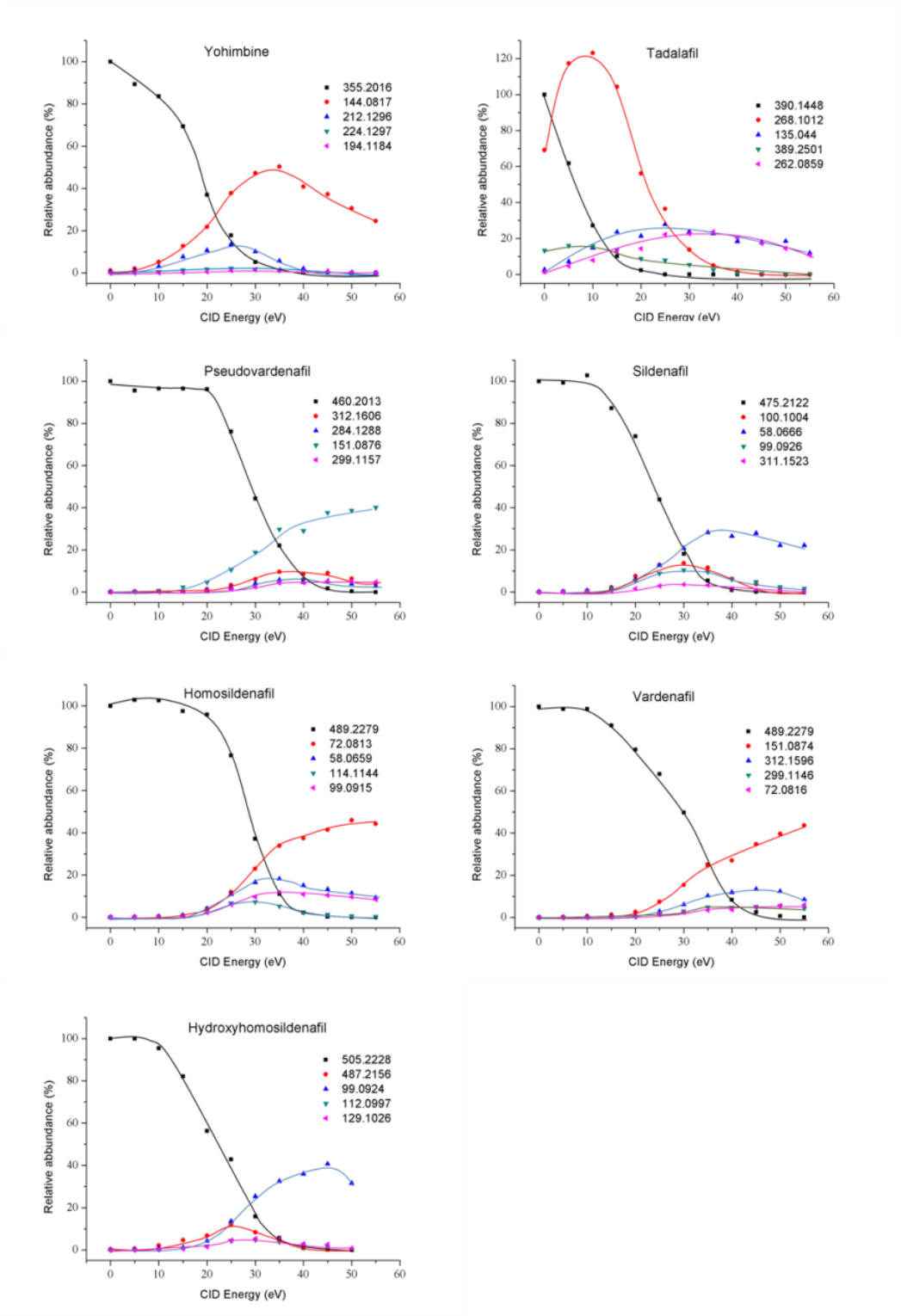

Figure 16. Relative daughter ion abundances *vs* collision energy (CE) energy for MS/MS experiments conducted with individual standard solutions at 100 ng/µL.

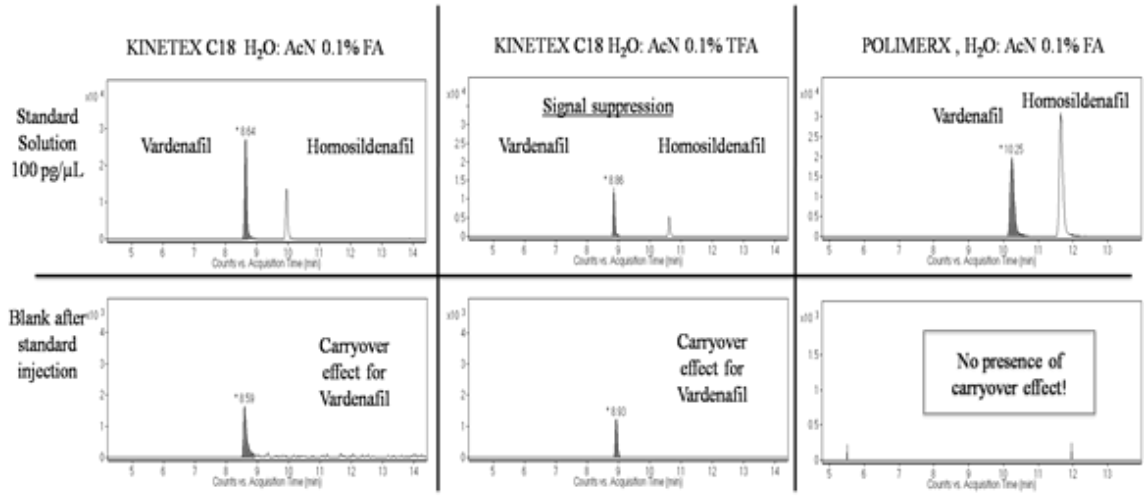

Figure 17. Chromatographic profiles for Vardenafil and Homosildenafil (isobaric, *m/z* 489.2279) using different columns and eluent modificants:

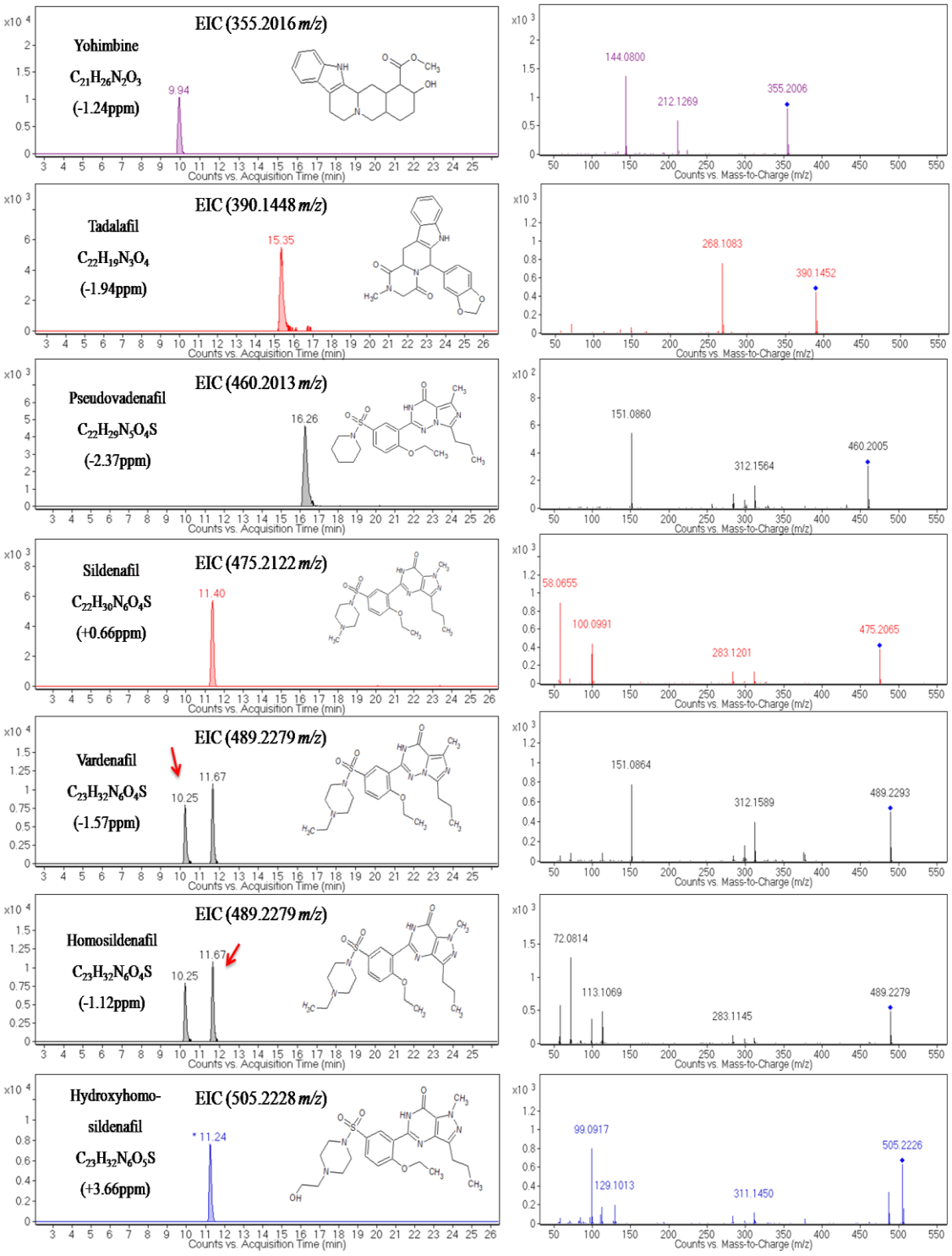

Figure 18. Chromatographic profiles and relative MS/MS spectra for target analytes at concentrations of 10 times LODs.

7.1.2 Optimization of the extraction conditions.

The sample preparation for trace analysis often involves several steps of clean-up, and enrichment with solid phase extraction. In the proposed method, the procedure is drastically simplified as it involves only a solvent extraction followed by sonication and centrifugation. The extraction parameters were tested in different conditions with a twofold purpose: at large concentration levels, typically encountered in the pharmacologic dosage, to cope with the counterfeit drugs, and at trace levels for analysis of cross contaminations.

Since certified reference materials were not available, the optimization of the extraction was made by using five commercial pharmaceutical formulations (see sample collection and preparation) of PDE-5 inhibitors, containing vardenafil, tadalafil, and sildenafil as API. Recoveries were calculated on the basis of the reported nominal amount of API and were expressed as concentration in the weighed matrix. The extracting solution (water, water:methanol 50:50, water:acetonitrile 50:50, all acidified with 0.1% FA), the sample amount (from 10 to 30 mg) and the extraction volume (from 1 to 2 mL) were optimized. Fig. 2 shows the percentage of recovery of the three synthetic API coming from the extraction of 10 mg of the relative milled tabs, varying the extraction solvent. The extraction of the most lipophilic tadalafil resulted very poor in acidified water, reaching the best yields with the water: acetonitrile solution. No difference in the extraction efficiency was observed for the different formulations of API. The optimization of the amount of the matrix and extraction volume was mainly evaluated in terms of efficiency and matrix effect. In ESI-MS technique, the matrix effect is described as the signal variation, mainly a suppression, of a compound in the ionization process due to the competition with matrix endogenous compounds. No further improvement of the extraction yields was obtained increasing the volume over 1.5 mL, whilst a significant matrix effect occurred when the amount of the sample extracted exceeded 15 mg (data not shown). Thus, the best conditions in terms of MME were obtained by extracting 10 mg of material with 1.5 mL of 50:50 water: acetonitrile solution acidified with 0.1% FA. However, when we tried to transfer this protocol for analysing bulk materials spiked at trace levels the same experimental conditions produced a quite relevant matrix effect. The slopes of the three-point calibration curves (1.0, 5.0 and 10.0 µg/g in matrix) obtained by spiking the five bulk materials after extraction using standard solutions have been compared. Results indicated that at low

concentrations, matrix components do interfere in the ESI ionization process, causing signal depletion, mainly for vardenafil and for the early eluted yohimbine in some bulk materials. Moreover, the extent of the matrix effect was sometimes dependent on the matrix composition. Extracts of pharmaceutical formulations did not produce that behaviour because the relative concentrations of the target analytes were much larger and taking into consideration the diluting factor. This drawback was anyway overcome using an extraction solution acidified with 1% FA, resulting both in a reduction of the ESI matrix effect and in the influence of the bulk composition. Therefore, it was possible to quantify the target analytes using a simple external calibration curve for all the considered herbal materials. Finally, no relevant changes in the extraction efficiency of the pharmaceutical formulations were experienced by using the extractant acidified with 1% FA instead of 0.1%. Figure 19 shows the percentage of recovery of the three synthetic API coming from the extraction of 10 mg of the relative milled tabs, varying the extraction solvent.

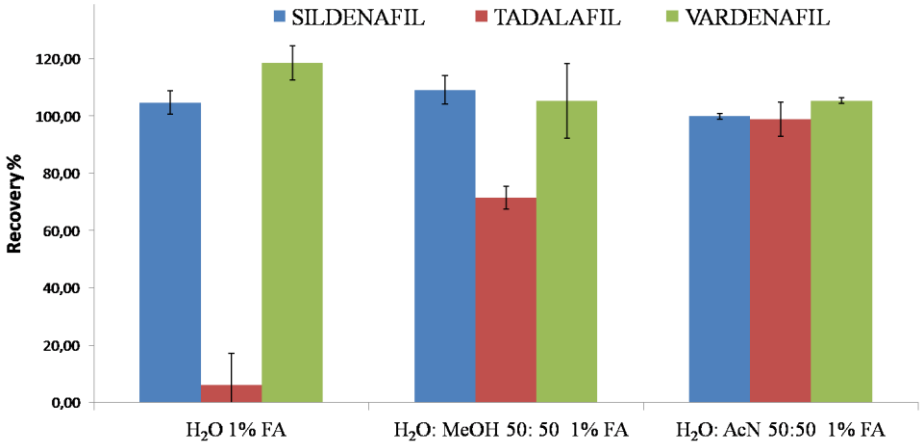

Figure 19. Recoveries and RSD for registered PDE-5 Inhibitors extracted from different pharmaceutical formulations with 1.5 mL of solvent. (n=6).

7.1.3 Validation.

The method selectivity was tested by comparison between five blank herbal matrices, not spiked and spiked at the lowest calibration level of the calibration curves. Mass accuracy obtained from the extracted currents of the quantifier $[M+H]^+$ ions were always better than 5 ppm without interferences of the matrix. Selectivity is also guaranteed by the presence of the qualifier ions for all the concentration levels in the fragmentation spectra. As previously reported, no significant matrix effect was

experimented in the optimized conditions, so that any consideration about linearity and LODs were made on conventional standard solutions.

In ESI-MS technique, the matrix effect is described as the signal variation, mainly a suppression, of a compound in the ionization process due to the competition with matrix endogenous compounds. Furthermore, a general matrix effect is generally experimented in the extraction procedure. In the present work, the possible presence of ESI-matrix effects was studied by comparing the regression slopes related to standard calibration and matrix-matched calibration procedures, obtained spiking blank samples of herbal bulk after extraction at the same concentration reported above. Comparable slopes within the statistical errors indicated that target analytes were free of matrix ionization suppression effect, so that quantification of target analytes can be made by using conventional external standard calibration.

The linearity for the proposed method was evaluated for each target compound by making a seven point calibration curve with injected amounts of 2, 5, 10, 50, 100, 500, 1000 pg (n=3), respectively. Calibration plots were made in terms of peak areas *vs* amount of injected analyte. The main parameters of the calibration curves were very satisfying (Table 10), with R^2 greater than 0.9991 for all target compounds and the relative residues always below 20%.

Table 10. Calibrations parameters and LODs for the proposed method.

Compound	calibration curve	R^2	LOD (pg injected)	LOQ (pg injected)
Yohimbine	y = 4794.2x + 5259	0.9999	1.04	3.11
Tadalafil	y = 372.5x - 598	0.9997	4.24	12.71
Pseudovardenafil	y = 6616.5x - 2867	0.9999	1.94	5.83
Sildenafil	y = 1808.2x - 198	0.9998	2.84	8.52
Vardenafil	y = 805.6x - 1548	0.9997	1.95	5.84
Homosildenafil	y = 2323.2x - 1920	0.9997	2.15	6.44
Hydroxyhomosildenafil	y = 1244.3x - 4332	0.9991	1.60	4.81

Limit of detection was rigorously calculated on the basis of a four-point calibration curve at low concentration (n=3) according to the Voigtmann method [87, 88]. As usual, LOQ was set = 3 LOD. LODs varying from 1 to 4 pg injected (corresponding to

30-120 ng/g in matrix) (Table 10) were considered satisfactory and suitable to detect target analytes even at trace levels.

Because of the unavailability of suitable CRMs, accuracy of the method was evaluated by spiking five different blank herbal formulations (Herbal powder formulation, gel for topic usage, and Herbal extract) at three concentration levels 1.0, 5.0, 10.0 µg/g of each target analytes (n=3, N=15). Trueness was computed as the mean value of thee replicates of each concentration level and ranged between 80.9% and 108.1% (Table 11). These values fulfil the criteria of 2002/657/EC and SANCO/10684/2009 guidelines [85, 86] that require recoveries in the range of 80–110% and 70–120%, respectively. Intraday repeatability was estimated by RDS values obtained from recoveries. Overall RSD values ranging from 2.7% and 10.8 % were considered very good, taking into consideration the variability of the five spiked matrices, and in accordance with validation guidelines.

Table 11. Accuracy of target analytes calculated from five different matrices spiked at three concentration levels (n=3, N=15).

compound	1.0 µg/g		5.0 µg/g		10.0 µg/g	
	Recovery (%)	Overall RSD (%)	Recovery (%)	Overall RSD (%)	Recovery (%)	Overall RSD (%)
Yohimbine	90.6%	10.1%	89.5%	8.7%	91.1%	10.8%
Tadalafil	101.6%	9.5%	106.5%	8.6%	92.6%	7.4%
Pseudovardenafil	91.0%	4.3%	96.1%	3.2%	93.6%	3.4%
Sildenafil	92.6%	4.7%	95.0%	10.7%	95.4%	9.1%
Vardenafil	94.8%	2.7%	80.9%	3.1%	87.4%	8.0%
Homosildenafil	97.7%	9.1%	99.7%	2.8%	93.1%	8.7%
Hydroxyhomosildenafil	108.1%	5.2%	102.1%	4.1%	100.3%	5.5%

7.2 Food supplement analysis

The proposed and validated method was applied to the analysis of 26 food supplement samples present in the Italian market. Two tablet samples resulted contaminated with sildenafil at the concentration barely above the respective LODs. EIC profiles for detected target analytes in experienced counterfeit samples with related MS/MS spectra are reported in Figure 21. The concentration levels are consistent with a cross contamination by authorized drugs, during storage or carriage, and are not able to cause

relevant adverse effects on human health. A capsule of a dietary supplement named Hero, bought in a web-store, was previously analysed using a LC-DAD system, resulting contaminated mainly by a compound with peak wavelengths at about 290 and 360 nm, characteristic of a thio derivative of Homosildenafil [31]. In 2008, FDA warned consumers this product containing an unapproved substance similar in chemical structure to sildenafil (http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2008/ucm116870.htm). At that time, that compound was not identify also because no certified standards were available for the quantitative analysis. Further analysis was made on two other capsules of the same lot only in the 2014 after the optimization of the presented LC-HRMS method. In this case, the MFE post run analysis showed two relevant signals identified as Homosildenafil (experimental *m/z* 489.2273, Δm=-1.2 ppm) or an its isomer, and one of the three isobaric compounds having a theoretical *m/z* 505.2050, namely thiodimethylsildenafil, thiomethisosildenafil and thiohomosildenafil. The signal of this second compound was two orders of magnitude larger than that obtained for hypothetic Homosildenafil. Anyway, although retention time of the supposed Homosildenafil, which standard was available, was within the query set, the observed fragmentation pattern in the AutoMS/MS acquisition did not correspond exactly. Thus, we found a compound in the web-free database *m/z* Cloud™, (https://www.mzcloud.org/) exhibiting the same fragmentation pattern and identified as the isobaric dimethylsildenafil (compound N°699 fragmented at 40 NCE in HCD mode). Analogously, the compound with experimental $[M+H]^+$ at *m/z* 505.2053 (Δm=+0.6 ppm) could be tentatively ascribed to both thiodimethylsildenafil (compound N°955) or thiomethisosildenafil (compound N°957) by comparing the available *m/z* Cloud™-MS/MS spectra registered at 40 NCE in HCD mode. Since the MS/MS fingerprints of the two compounds were unnoticeable, and although the acquired UV spectra was identical to that described for thiodimethylsildenafil [106], it was not possible discriminating among these two isomers. EIC profiles for detected target analytes in experienced counterfeit sample of Hero with related MS/MS spectra are reported in Figure 20.

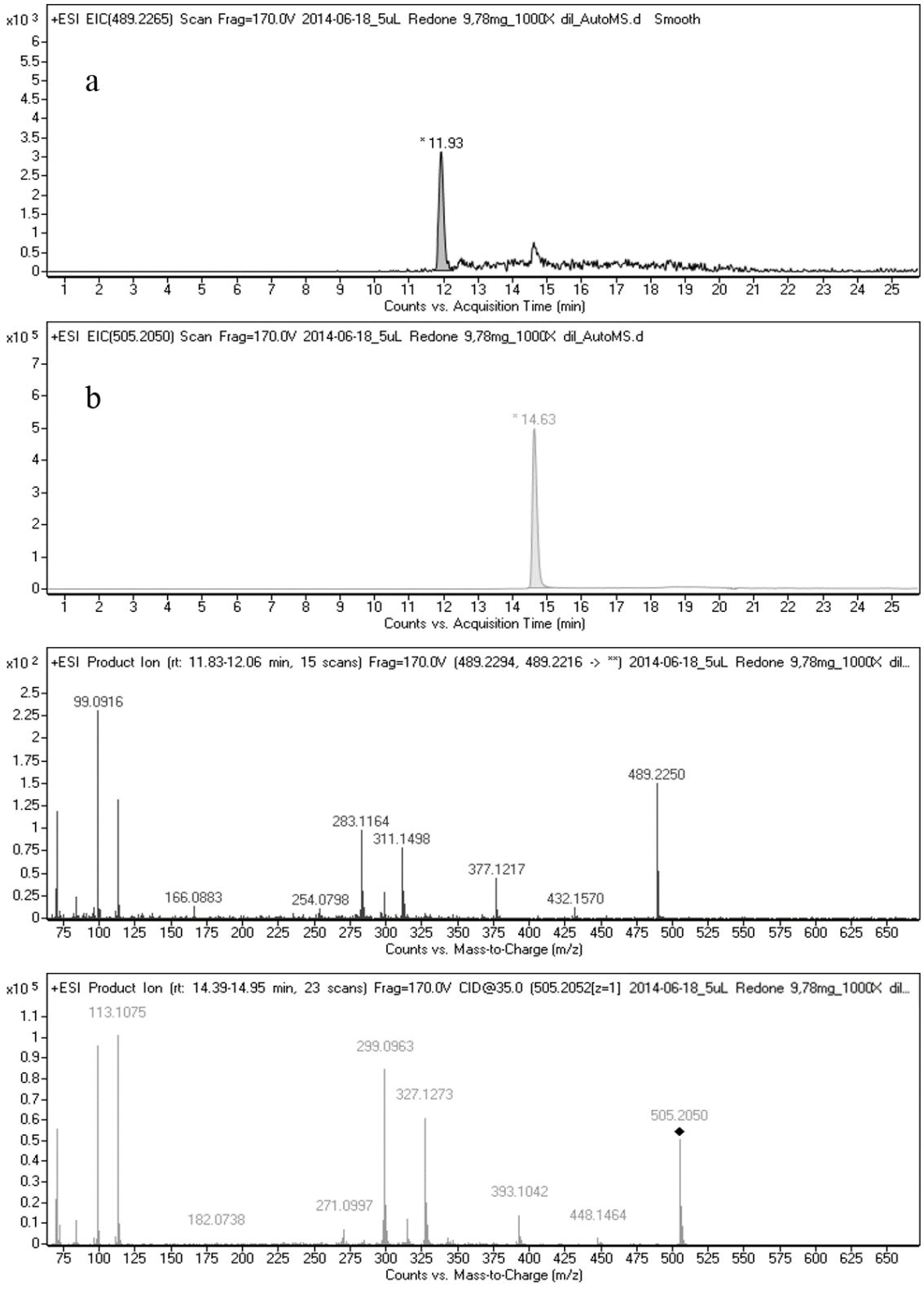

Figure 20. LC-MS chromatogram acquired for the analysis of a counterfeit sample of Hero with the related MS/MS spectra. EIC profiles of the compounds with theoretical *m/z* at a 489.2279 and b 505.2050.

After ascertaining that contamination was due to the inner bulk and not to the capsule shell, a semi-quantitation was made by diluting the extract by a factor 1000 like to the

authentic pharmaceutical formulations, and assuming the same molar response of the Homosildenafil. The assessed concentration for this single capsule is 0.25±0.02 µg/mg for dimethylsildenafil and 59.02±0.36 µg/mg for thiodimethylsildenafil or thiomethisosildenafil. The correlated dosage (0.11 mg/capsule of dimethylsildenafil + 25.12 mg/capsule of the thiodimethylsildenafil or thiomethisosildenafil), coherent with a pharmacological one and therefore may represent a concrete health risk for the unaware customers assuming this "natural herbal" food supplement. Finally, no other suspect PDE-5 inhibitor analogues resulted from data analysis made with the internal database, using the described MFE features.

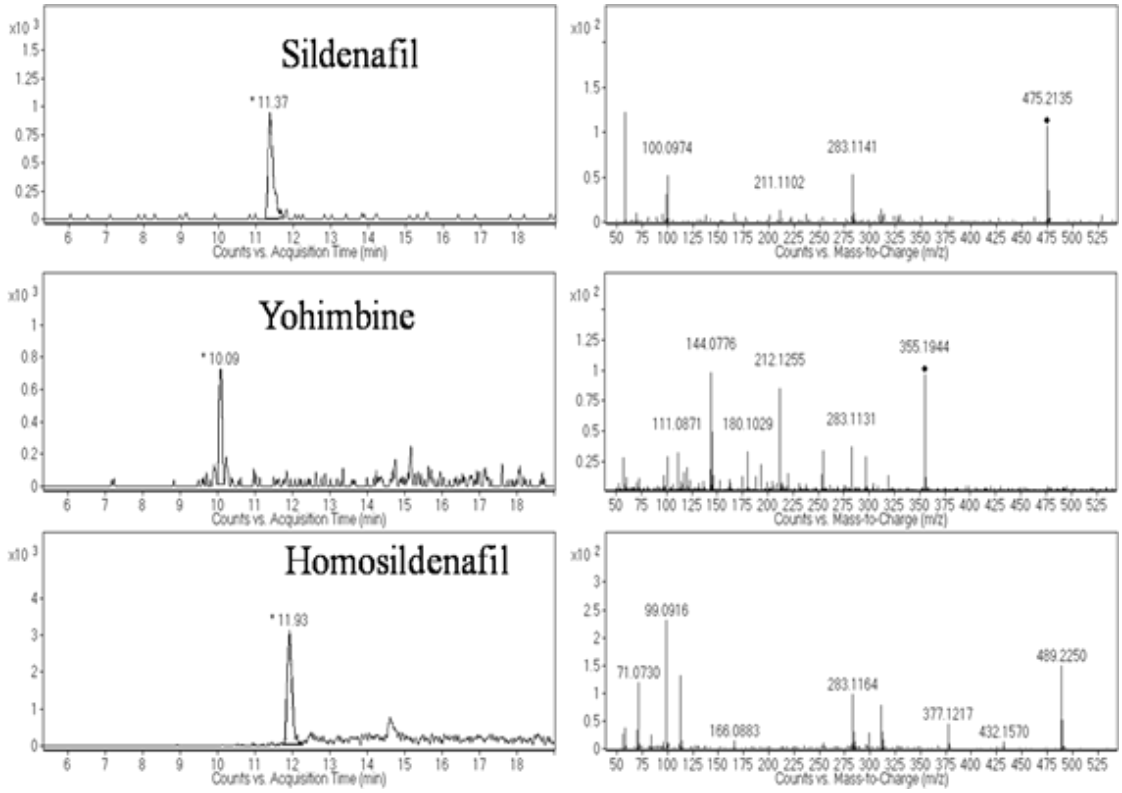

Figure 21. Chromatographic profiles for target analytes and relative MS/MS spectra in counterfeit samples.

7.3 **Conclusion**

The proposed method was proved to be very simple and robust, suitable for both screening and confirmatory purpose, allowing the identification of suspect PDE-5 analogues and the quantitation of the analytes with a single chromatographic run. The entire analytical process can be performed in less than one hour. A complete validation of the procedure was carried out, demonstrating a very high extraction efficiency, reproducibility and selectivity without incurring in any relevant matrix effect. The

performance of the extraction method makes possible both the determination of PDE-5 inhibitor analogues at high concentration for counterfeit analysis, and at trace levels for cross contaminations. The optimized method was successfully applied to the analysis of 26 real samples of natural dietary supplements and herbal remedies marketed for erectile dysfunctions. Three samples were found to be contaminated with synthetic PDE-5 inhibitors, both approved and unregistered. The possibility of further improvements of this procedure enlarging the database, even with fragmentation spectra, allows this method to be easily adopted by health agencies to contrast the illicit misuse of synthetic PDE-5 analogues.

8 NORMAN collaborative trial

8.1 Introduction

NORMAN network is a non-profit European association that through several initiatives would provide a help to the legislative organs in the field of environmental contaminants. It started its activities in September 2005 with the financial support of the European Commission.

The main objective are:

- Measurement methods harmonization for a better monitoring and risk assessment.
- Enhance the exchange of information and data on environmental emerging substances.
- Promote the maintaining and developing of knowledge of emerging pollutants stimulating interdisciplinary projects on problem-oriented research and knowledge transfer.

NORMAN network organizes many activities spacing throughout expert group meetings, workshops, databases and methods validation trials. As long-term goal, the association is also active in improving the identification of environmental unknown compounds and prioritizing of emerging substances. In August 2012, NORMAN started a successful cooperation with the web-mass library "MassBank" (http://www.massbank.jp) in order to fund a mass spectrum library focused on environmental pollutants. Coherently with the web vision, all the spectral information included in the database are addressed to improve the identification of unknowns and the access is free-of-charge.

In recent years, improvements of the analytical techniques have driven the interest of aquatic environmental scientist to the determination of organic pollutants at ultratrace levels. The scientific literature has depicted a scenario with a more and more number of compounds recognized in water, implying that the target analysis will be no longer sufficient to provide an exhaustive representation of the pollution status of the water bodies. On the other hand, the non-target analysis, needed to detect as many harmful substances as possible, is not harmonized and affected by a wide variability in the method implementation, making difficult the comparison of the obtained results.

In order to respond to this drawback, in 2013 NORMAN organized the first collaborative trial on non-target screening. The exercise involved many laboratories

across Europe (including the University of Padua) and contemplated the non-target analysis of the compounds present on a river water sample. A workshop has followed the trial with the aim of share the experiences, discuss the results, get to an agreement on harmonized terminology and workflows and, finally, get a proposal for further actions to be promoted in the field of non-target screening.

8.2 Trial results.

8.2.1 Target approach and suspect approach.

The analysis of the water extract was conducted using the "two step" protocol described in the section materials and methods. The column used was a Kinetex C18 (2.1X10 mm, 2.6 μm, Phenomenex) end the eluents were H_2O and AcN with 0.1% formic acid for positive mode and 0.1% NH_4OH for negative mode. The use of acidic and basic modificant was chosen to promote the ionization efficiently in the corresponding polarity. Injection of 40μL was perform to detect compounds at trace level.

A preliminary chromatographic run was performed in both positive and negative full scan MS, and the pseudomolecular *m/z* values of the available standards were used to extract EIC profiles with 20 ppm of mass tolerance. The presence of chromatographic peaks with significant S/N ratio in the EIC profiles was considered as a possible positive result.

At the moment of the collaborative trial, all compounds hade to be reported as target, suspect, non-target or unknown in a common excel spreadsheet (Table S 4) with all experimental and MS information useful for supporting the identification and comparing the results. This classification resulted vague when a comparison among different laboratories was made, and for the sake of simplicity the harmonized definition introduced in the Introduction chapter and proposed after the NORMAN workshop was here used.

The identification at level 1 was confirmed by comparison of retention times, and fragmentation spectra with the pure standards analysed in the same conditions. All the other non-confirmed possible positive results were subjected to MS/MS analysis, and then they were classified as non-target compounds with an identification confidence level from 2 to 5, depending on the structural information disused by the fragmentation spectra and on the proposal of a unique molecular formula.

Positive and confirmed target compounds were for perfluoroalkyl compounds, described as endocrine disruptors: perfluorooctanoic acid, perfluorooctansulfonic acid, perfuorohexanesulfonic acid and perfluorobutanesulfonic acid. As it was not possible to use surrogate standards nor to evaluate the matrix effect, a semiquantitation was attempted showing they were in the range of ng/L or sub-ng/L.

The suspect screening was approached on full scan analysis, searching for a limited number of compounds suggested by the NORMAN association (Figure 22) on the basis of previous monitoring campaigns, and employing libraries available for us: Pesticides (1600 entries), Forensic Tox (7300 entries), Synthesis (16,000 entries) from vendors, and in-house databases, *i.e.* cyanotoxins and PDE-5 inhibitors.

The libraries have been implemented in Molecular Features Extraction (MFE) setting the following parameters and thresholds: ion compound filters $\geq$ 1000 in MS level, and MME $\leq$ 20 ppm.

Positive results were carefully checked to eliminate false positive results due to:

a) ions recognized as salts adduct of the databases entries, but actually related to different signals;

b) ions clearly referable to series of homologues compounds, e.g. differing for an ethoxylate group;

c) ions belonging to the isotopic pattern of other signals;

d) ions ascribable to other compounds loosing water or ammonia.

f) signals close to dead time of the chromatographic run (t_R=0-2 min)

#	Compounds/Group	SMILES	CAS
Phosphates			
1.	Tributyl phosphate	CCCCOP(=O)(OCCCC)OCCCC	126-73-8
2.	Triphenyl phosphate	C1=CC=C(C=C1)OP(=O)(OC2=CC=CC=C2)OC3=CC=CC=C3	115-86-6
Phthalates			
3.	Diethyl Phthalate	CCOC(=O)C1=CC=CC=C1C(=O)OCC	84-66-2
4.	Dibutyl Phthalate	CCCCOC(=O)C1=CC=CC=C1C(=O)OCCCC	84-74-2
5.	DEHP	CCCCC(CC)COC(=O)C1=CC=CC=C1C(=O)OCC(CC)CCCC	117-81-7
Pharmaceuticals			
6.	Naproxen	C[C@@H](C1=CC2=C(C=C1)C=C(C=C2)OC)C(=O)O	22204-53-1
7.	Ibuprofen	CC(C)CC1=CC=C(C=C1)C(C)C(=O)O	15687-27-1
8.	Caffeine	CN1C=NC2=C1C(=O)N(C(=O)N2C)C	58-08-2
9.	Carbamazepine	C1=CC=C2C(=C1)C=CC3=CC=CC=C3N2C(=O)N	298-46-4
10.	Sulfamethoxazole	CC1=CC(=NO1)NS(=O)(=O)C2=CC=C(C=C2)N	723-46-6
Industrial Chemicals			
11.	Perfluoroheptanoate	FC(F)(C(F)(F)C([O-])=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	375-85-9
12.	Perfluorooctanoate	FC(F)(C(F)(F)C([O-])=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	335-67-1
13.	Perfluorooctan-sulfonate	C(S(=O)(=O)O)(F)(F)C(F)(F)C(C(C(C(C(C(F)(F)F)(F)F)(F)F)(F)F)(F)F)(F)F	29081-56-9
14.	Bisphenol A	CC(C)(C1=CC=C(C=C1)O)C2=CC=C(C=C2)O	80-05-7
15.	1-Decanamine, N,N-dimethyl-	N(CCCCCCCCCC)(C)C	1120-24-7
16.	Nonylphenol monocarboxylates	CCCCCCCCCc1ccc(cc1)OCC(=O)O	3115-49-9
Pesticides			
17.	Bentazone	CC(C)N1C(=O)C2=CC=CC=C2NS1(=O)=O	25057-89-0
18.	2,4-D	C1=CC(=C(C=C1Cl)Cl)OCC(=O)O	94-75-7
19.	Desethylatrazine	CC(C)NC1=NC(=NC(=N1)N)Cl	6190-65-4
20.	Atrazine	CC/N=C/1\N/C(=N/C(C)C)/NC(=N1)Cl	1912-24-9
21.	Isoproturon	CC(C)C1=CC=C(C=C1)NC(=O)N(C)C	34123-59-6
22.	Diuron	CN(C)C(=O)NC1=CC(=C(C=C1)Cl)Cl	330-54-1
23.	Terbutylazine	CCNC1=NC(=NC(=N1)Cl)NC(C)(C)C	5915-41-3
24.	Desethylterbutylazine	CC(C)(C)NC1=NC(=NC(=N1)N)Cl	30125-63-4

Figure 22. List of suspected substances provided for the collaborative trial according to previous surveys associated with the given sampling point.

Compounds unlikely to be present were deleted. The software outputted as sum of the suspect of the different libraries and polarity, 167 possible structures. Notable was the failing of the library "Synthetic" because of the low performance of the computer in processing such big amount of data, thus limiting our suspect screening capability.

The *m/z* values of the positive results were first used to extract EIC profile from the blank sample to manually check the presence of the same signal detected in the sample. Then, the *m/z* values converted in an inclusion list in AutoMS/MS experiments. The CID energy was set at 40eV, which was relatively high and ensured a high rate of fragmentation with several low mass fragments. A maximum number of MS/MS experiment per cycle of 8 and an exclusion time of 30 s was adopted to detect as many target as possible and ensure sensitivity and a good peak reconstruction.

The MS/MS spectra were evaluated by comparison with the spectra present in on-line web databases, especially MassBank and MzCloud. If sufficient structural evidences were provided by the fragmentation spectra, the substance was identified as level 2. Belonging to this class were 13 results (Table 12). When MS/MS spectra interpretation was not able to identify a unique structure, providing only information about the chemical class, the proposed identification was at level 3. Three structures were tentatively identified in this way: dihydroxy-octadecenoic acid, hydroxy-octadecenoic acid and dihydroxy-linoleic acid. For these compounds, despite the presence of characteristic fragments identifying the functional groups, the relative position of the hydroxylation was not inferable by the MS/MS analysis.

Finally, positive results whose assignments were not confirmed by MS/MS, were declassified as non-target compound and treated with a different approach.

Table 12. Number of detected compounds divided by screening approach adopted and identification level.

	level 1	level 2	level 3	level 4	level 5
target	4				
suspect		13	3		
non-target			17	6	57

8.2.2 Non-target approach

Non-target analysis was primary focused on the best peak chromatogram (BPC) in which the profile is the current ion of the most intense signals. This approach was chosen for the manually detection of the ions present. From the MS spectra of the chromatography peaks, the most intense signals were selected and, after a raw evaluation of the isotopic pattern, for each ones:

a) The molecular formula was generated, setting the software algorithm with a limited number of carbon, hydrogen, oxygen, nitrogen, sulphur, chlorine and bromine.
b) The *m/z* value was input in an inclusion list to perform the target MS/MS experiments.

Target MS/MS was conducted setting the same parameters as suspect screening. The majority of non-target results were at level 5 (57 mass of interest) because the formula generated had an non acceptable high mass error or was likely to be non-natural.

Twenty-one unique molecular formulas were successful characterized and for 17 of them a tentative identification was proposed at level 3 by evidences arising by MS/MS experiments. Some tentative identifications were achieved taking advantage of the retention time standardization through the RTI, and comparing results with the RTI database implemented by some researchers belonging to the NORMAN group.

8.2.3 Retention time index (RTI)

An important objective of the trial was to assess the use of retention time information in the LC screening approaches.

Table 13. Retention times for the analyte for the RTI calculation in positive and negative modes.

substance name	Monoisotopic MW (Da)	logP	Tr (min) Positive detection	Tr (min) Negative detection
Metformin	129.10143	-1.36	1.1	
Chloridazon	221.03558	1.11	10.6	9.5
Carbetamide	236.11608	1.65	12.5	11.28
Monuron	198.05598	1.93	13.0	11.86
Metobromuron	258.00043	2.24	15.4	14.18
Chlorbromuron	291.96146	2.85	17.3	16.06
Metconazole	319.96146	3.59	18.5	
Diazinon	304.10104	4.19	19.9	
Quinoxyfen	306.99669	4.98	20.9	
Fenofibrate	360.11282	5.28	22.4	

We measured the substances mixture provided by the NORMAN foundation in the described experimental conditions in both the ESI polarities. Table 13 reports the detected compounds and respective retention times. It is possible to note that the majority of substances used were ionized only in positive mode, while in negative mode only five of these have been efficiently ionized. The retention times were plotted against the logP (Figure 23) of the respective compounds shared within the NORMAN material, and the linear fitting of the data was done. The equation obtained was used to calculate a retention time index (RTI) for the LC compounds identified. These data was used according the instruction furnished in the RTI database, which performance was tested in the frame of this collaborative trial. The database collects some thousands of environmental substances with RTI, CAS number, structures, commercial use and other information. On the basis of the experimental chromatographic conditions used and RTI

inserted, the database propose some compounds listed with the respective MS accuracies, which have to be anyway evaluated in terms of MS/MS fragmentation (Table S 4)

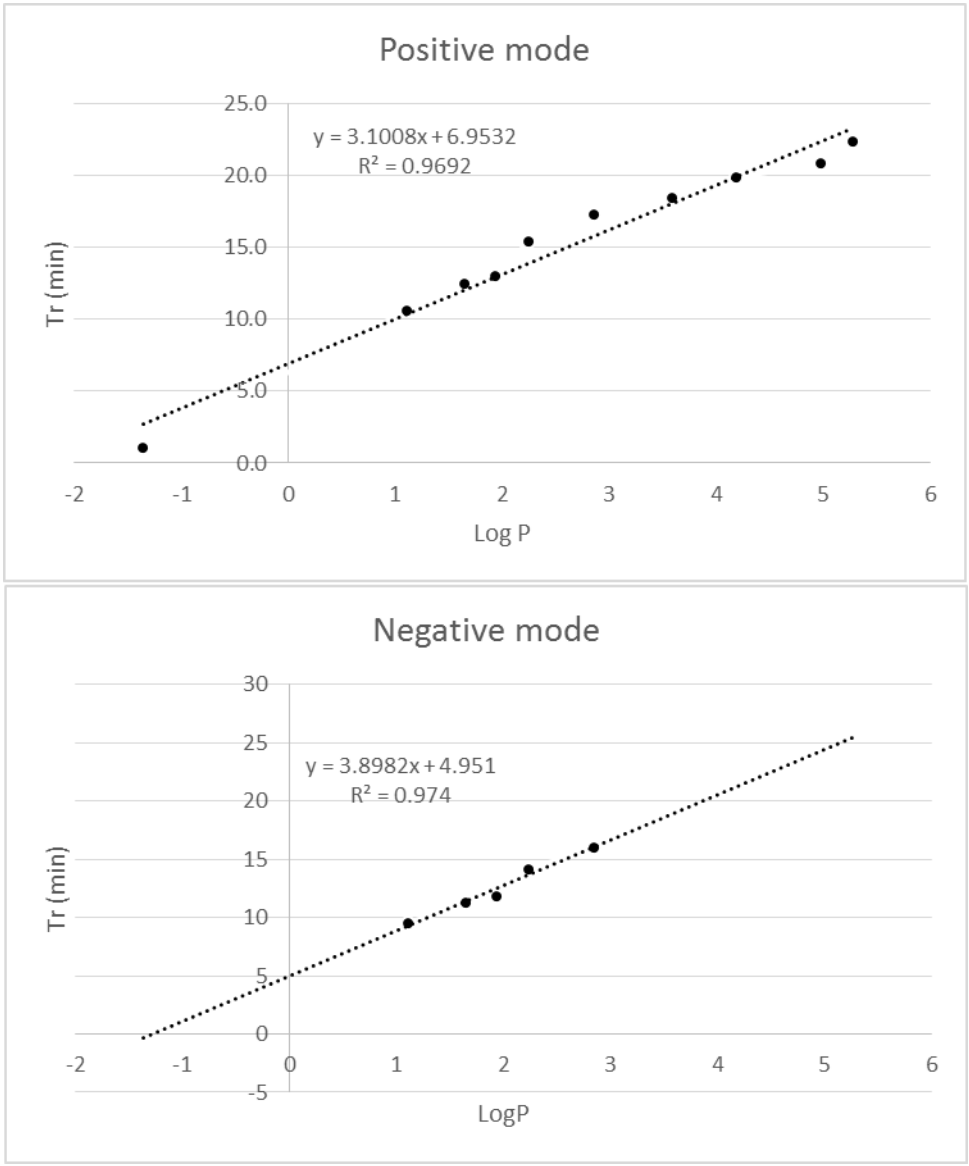

Figure 23. RTI calibration curve for positive and negative mode and linear fitting (dotted line).

8.3 Trial consideration and Conclusions

The laboratories participants to the trial had different backgrounds; either experienced in non-target analysis either novice in the field, attending the screening analysis for the first time. For our group was in absolute the first attempt in such comprehensive non-target analysis, and this fact was reflected in the relative low number of submitted results of non-target compound.

The used instrumentation setting was roughly comparable with the rest of the trial's participants (Table S 2 and Table S 3). The adopted "two-step" protocol was excellent for the target analysis. In fact, this approach allowed a fast, easy and reliable confirmation of the structure. The unique limitation of the target screening we have experimented was the very limited availability of analytical standards.

For the suspect screening the limitation was the non-availability of a database containing specific water contaminants, so that we used vendor's libraries in which most of the compounds were non likely to be present in river water. Structural investigation using target MS/MS feature was resulted critic to increase the identification level of the structure. Q-TOF system is known for the speed in acquisition and this allows the acquisition of many MS/MS spectra at the same time. Despite this, a partial sensitivity loss was experimented with the increasing number of fragmentation per cycle, thus limiting the overall capability in structural confirmation, especially when analytes were at ultratrace levels. To avoid this drawback, similarly to all the trial participants, we adopted a quite long chromatographic run in order to give a sufficient time to the analyser for performing the MS/MS experiments and having a good peak reconstruction in EIC spectra.

In non-target analysis, in addition to the same issues as suspect screening, the generation of a unique and reliable molecular formula seems to be the limiting step. As expressed by most of the participants, non-target analysis require specific software that are non-commercially available able to treat big dataset and produce reliable results. Finally, the non-target screening was tedious and highly time-consuming because, without any previous experience in the field, and the non-availability of specific software required a big amount of manual work. The needing of much more time for the data treating was a common feeling of the trial participants, suggesting that it is currently impossible adopting such non-target approach for the routinely analysis of the water contaminants, although a rough harmonization of the analytical protocol could be certainly indicated.

9 Development of a workflow for HRMS analysis of $PM_{2.5}$ organic fraction: post-run data analysis and the role of ionization sources.

9.1 APPI analysis optimization.

9.1.1 Source parameters optimization

Initial optimization of the APPI measures was focused on the source parameters using a standard mixture of PAHs, Nitro-PAHs and O-PAHs in a methanol:dichloromethane 1:1 solution. The concentration levels (Table 4) were in the range 6-133µg/mL for PAHs, 0.6-5.3µg/mL for Nitro-PAHs and 0.13-13 µg/mL for O-PAHs.

The standard mixture was analysed in direct infusion at different temperatures from 50 to 350ºC setting the mass range to *m/z* 100-650 and the flow at 10 µL/min. For each temperature, after the stabilization of conditions, the mass spectra were recorded for 30 seconds. For each measure, the overall intensity and spray stability were assessed through the average of the single TIC values and its standard deviation respectively. As Figure 24 shows, with the increasing of the source temperature both TIC values and related variances increased. More specifically, looking at the responses of single analytes of the mixture (data not reported), for two nitro-PAHs (4-nitrocatechol and 4-nitrophenol) and one O-PAH (4-phenanthrenecarboxaldehyde) a decreasing in intensity was experimented over 200-250°C, indicating a thermal decomposition (Figure 25). The spray stability dramatically decreased over 200°C as represented by the standard deviations values. Thus, a temperature of 200°C was chosen for the APPI analysis in order to prevent the loss of Nitro and oxidized PAH, having and at the same time a good overall sensitivity and spray stability.

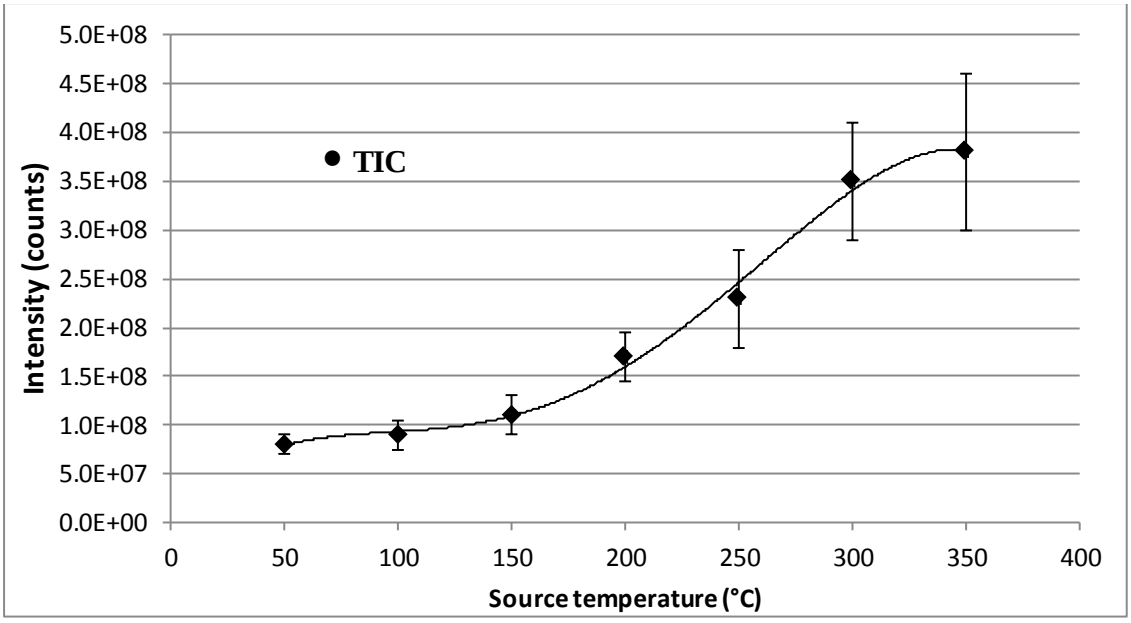

Figure 24. TIC intensity trend with the APPI source temperature.

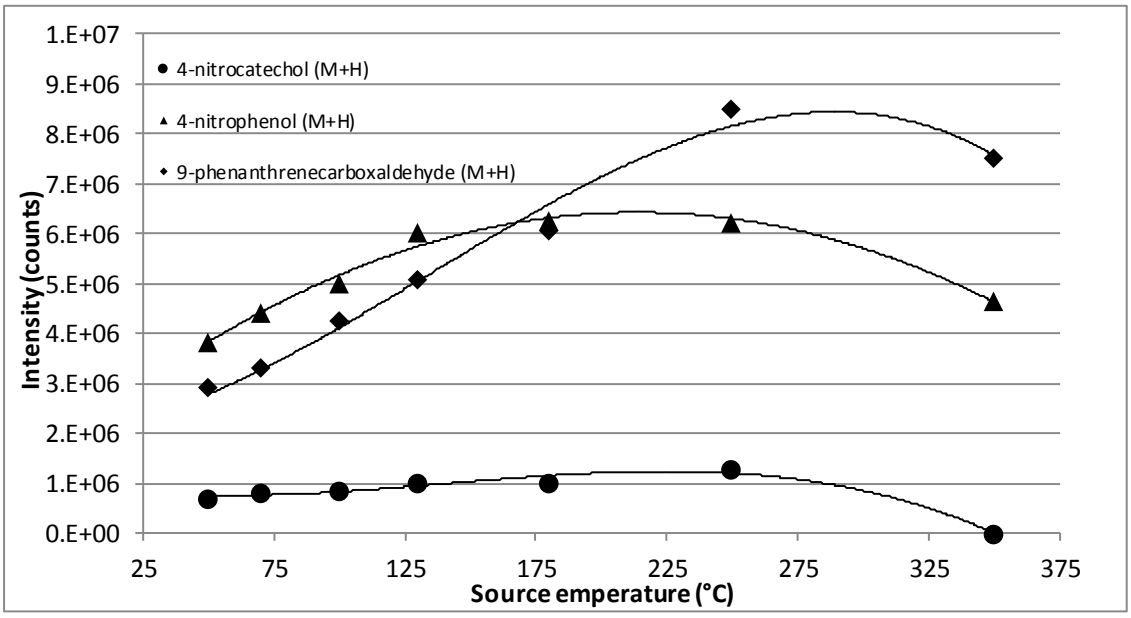

Figure 25. EIC intensity trend of selected nitro-PAHs and O-PAHs with the APPI source temperature.

An analogue procedure was followed for the gas flows optimization. The histogram plot in

Figure 26 shows the TIC average values with standard deviations for each tested pair of auxiliary gas and sweep gas. Flow rate of 5 and 10 arbitrary units for auxiliary gas and sweep gas were chosen respectively because representing a good compromise between response and spray stability.

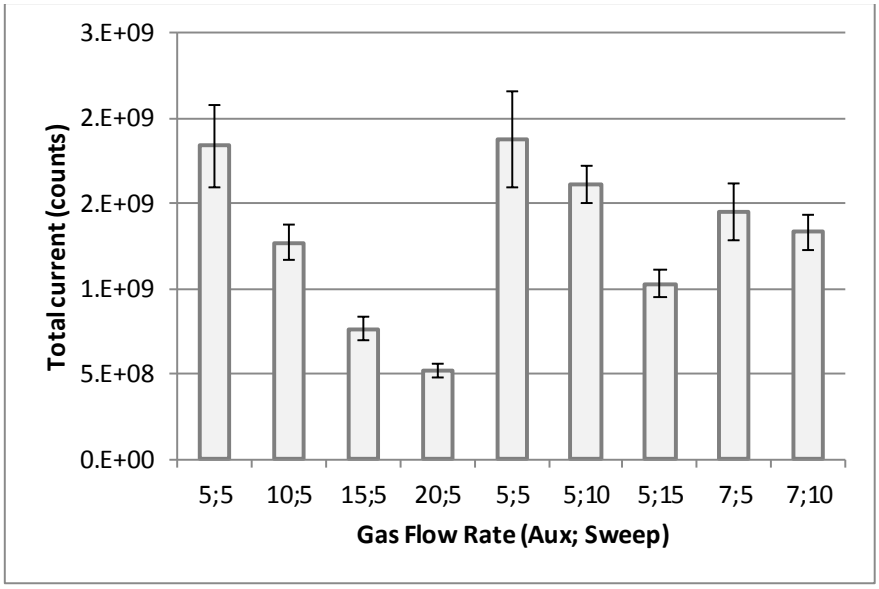

Figure 26. TIC intensity and spray stability (bars= means ± SD) using different combination of auxiliary and sweep gas.

9.1.2 Dopant optimization

Toluene and acetone were tested as dopant agents at concentration of 5 and 10% (v/v) in the mobile phase for improving APPI ionization efficiency. The best results were showed by using toluene at 10%. A complete table of intensities is reported in the appendices (Table S 6), and the Table 14 reports the sums of the intensities of the $[M+H]^+$ and $[M^{\cdot}]^{+\bullet}$ ions and the relative percentage with respect to the higher signal registered. As general observations:

- Toluene at concentration of 10% produced the best results in terms of intensity of the standard's signals for all the class compounds.
- PAHs formed manly molecular ions, while Nitro-PAHs and O-PAHs preferentially generated quasimolecular ions $[M+H]^+$. 4-nitrocatechol and 4-nitrophenol formed only quasimolecular ions (Figure 27).
- 4-nitrocatechol signal was registered only using toluene, with the best results experienced with 10% toluene with respect to 5% (about an half of the signal), whilst acetone never produced an appreciable signal.
- Acetone promoted the formation of quasimolecular ions, increasing the ratios between quasimolecular and molecular ion, as showed in the Figure 27.

Table 14. Sum of the EIC intensities for the chemical classes and all the substances, analysed using different solvent dopants with relative percentage with respect to the greater value reported in parenthesis.

	Compound class	Toluene 10%	Toluene 5%	Acetone 10%	Acetone 5%
Quasimolecular ions (M+H+)	PAH	$1.08*10^8$ 100%	$7.66*10^7$ 71%	$1.04*10^7$ 10%	$1.28*10^7$ 12%
	Nitro-PAH	$1.23*10^7$ 100%	$7.72*10^6$ 63%	$3.21*10^5$ 3%	$4.80*10^5$ 4%
	O-PAH	$1.34*10^8$ 100%	$8.55*10^7$ 64%	$6.29*10^7$ 47%	$6.26*10^7$ 47%
	Total	$2.54*10^8$ 100%	$1.70*10^8$ 67%	$7.36*10^7$ 29%	$7.58*10^7$ 30%
Molecular ions (M+)	PAH	$6.56*10^8$ 100%	$3.79*10^8$ 58%	$3.11*10^7$ 5%	$3.65*10^7$ 6%
	Nitro-PAH	$1.26*10^7$ 100%	$5.18*10^6$ 41%	$4.39*10^5$ 3%	$5.97*10^5$ 5%
	O-PAH	$3.27*10^6$ 100%	$1.58*10^6$ 48%	$9.84*10^5$ 30%	$9.09*10^5$ 28%
	Total	$6.72*10^8$ 100%	$3.86*10^8$ 57%	$3.26*10^7$ 5%	$3.80*10^7$ 6%
All Ions		$9.26*10^8$ 100%	$5.56*10^8$ 60%	$1.06*10^8$ 11%	$1.14*10^8$ 12%

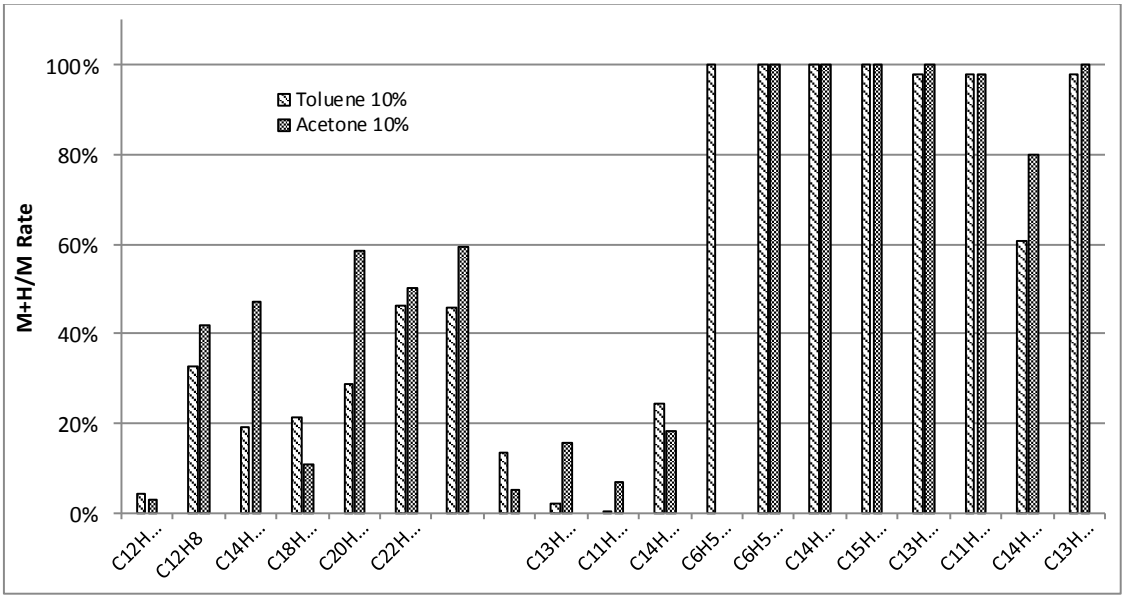

Figure 27. Intensity ratio between quasimolecular $[M+H]^+$ and molecular ions $[M]^{+\cdot}$ using toluene 10% and acetone 10%

9.1.3 Recovery study on spiked blank filters.

The study of the recovery for the extraction procedure was carried out spiking Teflon blank filters at concentrations of PAHs close to those expected in real samples. As reported in literature, the average winter concentration of the sum of PAHs in north Italy is in the range 20-50 ng*m^{-3} and Benzo[α]pyrene (B(α)P) represent the 17% of the total amount (3.6-8.5 ng m^{-3}).

Two pieces of 1/8 of filter were spiked with 20 µL of the stock solution and then one was extracted using 5X15 mL of methanol and the other using the same volume of methanol:dichloromethane 1:1. The 15 mL of extracts were evaporated under a gentle nitrogen flow until a final volume of about 3 mL. The obtained extracts were filtered using 0.22 µm Teflon filter and divided in two aliquots of 1.5 mL. For each aliquot, extract one aliquot is evaporated to dryness and then reconstructed using 1.5 mL of the same solvent. The standard solution was obtained by dilution of 20 µL in a final volume of 3 of the corresponding solvent.

The Table S 7 reports all the intensities of the detected ions obtained with the different extraction solvents, while Table 15 summarizes the experimental highlights.

Table 15.Recovery study results from blank samples spiked with PAHs, nitro-PAHs and O-PAHs standard solution at reported in Table S 7.

Number of detected Ions					
St. Mix Methanol	Spiked sample Methanol	Spiked sample Methanol reconstituted	St. Mix Methanol:CH_2Cl_2	Spiked sample Methanol:CH_2Cl_2	Spiked sample Methanol: CH_2Cl_2 reconstituted
26	29	21	28	25	21
average values of recoveries					
Ratio St. Mix Methanol/ CH_2Cl_2: Methanol	Recovery Methanol extraction	Recovery CH_2Cl_2: Methanol extraction	Recovery Methanol reconstituted extraction	Recovery Methanol: CH_2Cl_2 reconstituted extraction	Ratio Sample extraction Methanol/ CH_2Cl_2:Methanol
78%	105%	57%	47%	31%	122%

Considering at first instance the number of compounds found in the mass spectra of the different sample preparations, the extraction using methanol allowed to recognize a greater numbers of ions with respect of the extraction using the mixture methanol:dichloromethane 1:1 (29 instead 28). Although the standard solution gave a

better response in methanol:dichloromethane 1:1, the mean recoveries was higher for the methanol extract.

The same extraction conditions were used on aliquots of one quarter of filter from a winter sample (code FP-1) and analysed both in positive and negative modes in order to assess the best protocol in terms of number of molecular formulas obtained from data analysis, as common ions of triplicate measures (Figure 28). Among the three organic extractants evaluated, methanol showed the highest performance in terms of ability to elute compounds generating a larger number of molecular formulas. These results can be explained by a better solubility of the compounds present in the aerosols.

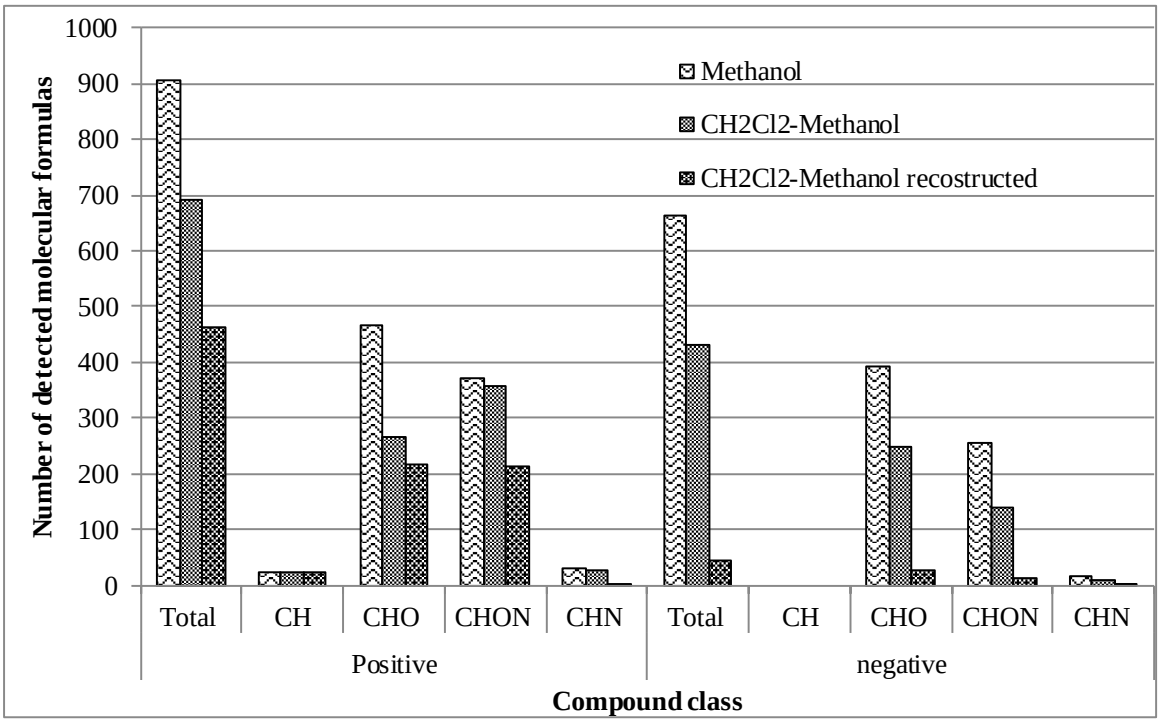

Figure 28. Number of inferred molecular formulas using different extractant solution.

9.2 Algorithm development

9.2.1 Previous algorithms and criticisms

Two different algorithms for positive and negative nanoESI molecular formula filtering were previously developed in the centre of Atmospheric Science group of the University of Cambridge (http://www.atm.ch.cam.ac.uk/). As inputs of the algorithms, a table containing the name of the sample, its mass drift range, noise level and "signal to noise" had to be provided. The mass error ranges were calculated looking for at least five known contaminants present in the MS measures, taking their mass errors and adding and subtracting 0.5 ppm respectively to the maximum and minimum value of the set, obtaining in this way the "lowerppmLimit" and "upperppmLimit".

The noise was calculated sampling it manually in three different regions of the mass spectra, calculating the average and standard deviation and finally adding three times the standard deviation to the mean. The "signaltonoise" value was used for discriminating signals to be ascribed to samples from blanks, and entries with values below five was deleted.

The original algorithm performed the following steps both on the sample and on the blank:

1. Importation of the raw .csv table and addition of 17 columns named: C; H; N, S, O, 13C, 34S, Na, Theo n.Mass, DBE, N comp, C, H, N, S, O, ^{13}C, ^{34}S. The new table formed has 25 columns and it was called xcalraw.
2. The algorithm for each row takes the molecular formulas, that have the general structure of CHNSONa, and splits them writing in the corresponding columns the number each elements present in the formula.
3. The algorithm selected only the molecular formulas with: mass errors >ppm-lower; mass error <ppm-upper;$^{12}C+^{13}C$ > 0; 4) H > 1; H(+1 if Na adduct)/($^{12}C+^{13}C$) < H/Cmax; H(+1 if Na adduct)/($^{12}C+^{13}C$) > H/Cmin; N/($^{12}C+^{13}C$) < N/Cmax; N/($^{12}C+^{13}C$) > N/Cmin; O/($^{12}C+^{13}C$) <O/Cmax; O/($^{12}C+^{13}C$) >O/Cmin; S/($C+^{13}C$) <S/Cmax.
4. The algorithm copied the elemental composition in the columns corresponding to the neutral molecular formulas. For the concerning of hydrogen the number is decreased by 1 or in the case of Na is present no subtraction is performed. The model considers all the ions M+H or M+Na.

5. The algorithm calculated the DBE number for all the neutral molecular formulas starting from the compositions.
6. The algorithm calculated the exact molecular mass for all the neutral molecular formulas starting from the compositions.
7. The algorithm selected only the rows where an integer number of DBE is present for the neutral molecular formulas.
8. The algorithm sorted the rows in order of decreasing intensity (column 2)
9. The algorithm selected only the rows where the formula($C+^{13}C$)*12 + H + N*14 + ($^{32}S+^{34}S$)*34 + O*16 + N is giving an even result.
10. The algorithm divided the rows in four tables contain each one the rows where are present; a) both ^{12}C and ^{32}S; b) only ^{13}C or ^{34}S; c) only ^{13}C; d) only ^{34}S. The code started a loop checking the intensity ratio between the isotopes. The only row passing the constrictions were selected and subtracted to the previous table. The output-table contained all formulas except the isotopes recognized.
11. The algorithm selected only the molecular formulas where no ^{13}C or ^{34}S are present.
12. Only rows with intensity higher than noise were selected. This step was performed on the sample and on the blank using the same noise value.
13. The algorithm subtracted the molecular formulas of the blank from the ones of the sample and performed the intensity check in accordance with the value "signaltonoise".
14. The algorithm exported the list of filtered molecular formulas.

The algorithms, although easy to be implemented, were affected by some criticisms and conceptual faults. First, the algorithms were suitable only for MS signals typical of the ESI source, arising only from deprotonation ($[M-H]^-$, negative mode) or protonation ($[M+H]^+$, positive mode). Therefore, it was not adapt to APPI analysis, where a probable ionization behaviour is the formation of radical molecular ions ($[M]^{\bullet+}$ or $[M]^{\bullet-}$).

Indeed, the algorithm subtracted each mass of the obtained neutral molecular formula from each blank sample; when the results was 0 (same compound), the algorithm checked the intensity ratio deleting the entry if did not exceed a value of five. The selection of the molecular formulas was made within the mass error range calculated on known compounds present in the mass spectra, adding a further tolerance of 0.5 ppm.

Although this is an easy approach, it was affected by the presence of false-positive results. The possible and most frequent related scenarios were:

1) Same *m/z* values in real and blank samples with different bias, but inside the error range covered by the assigned formulas: although the formulas assigned are the same, the mass error filter selects different compounds in the two samples, leaving the ones selected in the real sample.
2) Different *m/z* values in real and blank samples with equal or different bias inside the error range covered by the assigned formulas. Two cases of false positive results are possible: it can be possible to have molecular formulas associated to the peaks different between real and blank samples, or the same assigned molecular formulas with different mass error, as in the previous case.
3) Different *m/z* values and bias larger than the error range measured by the assigned formulas. The mass error filter selects all the formulas in both blank and sample, and in the case that the assigned formulas are different (more frequent at high *m/z* values), no subtraction is performed.

Third, the noise definition was considering only the first distribution of noise signals while another mode was present at greater intensity (see below). This can potentially introduce false-assignment in the final molecular formula list and anyway lengthened the analysis time.

Finally, the subtraction of blanks as the last step of the algorithm required long time, due to the large amount of operations to carry out.

9.2.2 Algorithm modifications

9.2.2.1 Data analysis via "MassSpecProcessing v1.0" algorithm

In order to select to most chemically realistic molecular formula among those proposed by Xcalibur™, the CSV lists were processed with the algorithm "MassSpecProcessing v1.0". This algorithm was an evolution of the previous algorithm able to process data obtained in nanoESI and APPI in positive and negative polarity. The input parameters and the structure of the code were drastically modified maintaining their basic heuristic rules. "MassSpecProcessing v1.0" has as inputs the name of the folder contain all the spectrums in CSV-file and an input table containing the parameters reported in the Table 16.

In this algorithm, also an elemental ratio constriction have to be included in the specific section. The values were selected from statistic studies on compound libraries with the aim of deleting the unlikely natural compounds. These values were maintained constant and were H/Cmin 0.3; H/Cmax 2.5; O/C max 2; O/Cmin 0; N/Cmax 1.3; S/Cmax 0.8.

Table 16. Input parameters for "MassSpecProcessing v1.0".

Parameter name	Description
Sample Name	Identifies the csv-file relative to the sample
sample mass drifts	Estimator of the mass accuracy for the sample measure (μ_S)
sample SD	Estimator of the mass precision for the sample measure
N Sample	Number of compound used to SD calculation for the sample
Sample mass drift filter factor	A multiplicative factor of sample's SD for preselection of the molecular formulas on the basis of mass accuracy
Sample noise	Intensity value equals to three time the mean value of the noise distribution of the sample mass spectrum
Blank name	Identifies the csv-file relative to the blank
Blank mass drift	Estimator of the mass accuracy for the sample measure
Blank SD	Estimator of the mass precision for the sample measure
N Blank	Number of compound used to SD calculation for the blank
Blank noise	Intensity value equals to three time the mean value of the noise distribution of the blank mass spectrum
Sample to noise ratio	The minimum value that the ratio between a signal in the sample and the same signal in the blank (when detected) to not delete the sample entry in the blank subtraction
Acquisition mode	Abbreviation of the ion source and polarity used in the acquisition of the mass spectrum and could be: APPI NEG, APPI POS, ESI NEG, ESI POS

The sequence of operations performed by this algorithm was:

1. Noise subtraction: all the peaks with intensities below the noise level (NL) were removed. This step was performed both for the sample and the blank on the basis of their own NLs.
2. Blank subtraction: each signal was considered positive only when exceeds of five times the intensity of the corresponding signal in the blank, if present.
3. Transposition of the molecular formula in elemental composition.
4. Preselection of molecular formulas within the error range determined by the mass accuracy and mass precision.

5. Filters of molecular formulas not compliant with the constrictions on elemental ratios.
6. Determination of the neutral molecular formulas taking into account the variability of the number of hydrogens with the acquisition mode input.
7. Calculation of DBE, exact neutral molecular mass, oxygen to carbon ratio (O/C), hydrogen to carbon ratio (H/O), oxidation state for the carbon (OSc), Kendrick mass (KM) and Kendrick mass defects (KMD) for the neutral molecular formulas.
8. Removal of the neutral molecular formula showing a non-integer value of DBE.
9. Removal of the neutral molecular formula non-respecting the nitrogen rule.
10. Removal of signals imputed to isotopic contributions (^{13}C and ^{34}S).
11. Selection of the most accurate molecular formula.

For the concerning of steps 5 and 6 a switching to different scripts was introduced in the algorithm in order to taking into account the different ionization pathways that can be generated from the different sources and polarities employed.

Including ^{13}C and ^{34}S isotopes is not chemically important, but their inclusion in the assignment processes increased the reliability of the other assignments. However, the inclusion of more isotopes enlarged the number of multiple formula proposed by the software. In the step 6, algorithm detects the presence of the isotopic patterns. It is assumed that each formula containing isotopes must have the corresponding monoisotopic equivalent with higher intensity. When the isotopic patterns are detected the assignments containing ^{12}C o ^{32}S are confirmed, while assignments derived by the isotope are removed by the data treatment process.

9.2.2.2 APPI compatibility

In order to taking into account the possible presence of the molecular ions in the mass spectra acquired using APPI source, the calculation of the number of hydrogens in the neutral formulas has been changed. For negative APPI mode the new algorithm considers raw formulas showing integer DBE values as molecular ions $[M]^{\bullet -}$, while those with non-integer DBE values are classified as deprotonated quasimolecular ions $[M-H]^-$. In the same way, for positive APPI mode, ions showing a non-integer value of DBE were considers quasimolecular ions $[M+H]^+$, and ions with integer DBE molecular radicals $[M]^{\bullet +}$. Compounds showing both molecular and quasimolecular ions were reported in the algorithm outputs associated at the same neutral molecular formula. The

algorithm was programmed to apply different ionization rules to the measures obtained with different ionization source. All ions produced in negative nanoESI were considered deprotonated ($[M-H]^-$), ions produced in positive nanoESI were considered protonated ($[M+H]^+$) or sodium adducts ($[M+Na]^+$) on the basis of the presence or not of sodium in the ion molecular formula, ions produced in negative photoionization were treated as described above.

9.2.2.3 Noise subtraction

In the developed algorithm, the noise subtraction was performed at the beginning of the process in order to have into a limited number of data, thus drastically decreasing the elaboration time.

Previously, the signal to noise is calculated manually sampling the noise, calculating the average and its standard deviation and adding three times the latter to the main value of the noise. A first attend was focused on the application of the rigorous definition and was quite easily to be implemented, sampling the noise in the characteristic *m/z* gaps (signals are present primarily nearest the unitary values) present in the ordinary mass spectra as show in the Figure 29. Although this procedure was well characterizing the main noise distribution, concrete statistical issue affected this approach. Looking at the Figure 30 and Figure 31, showing the signal frequency *vs* the MS signal intensity, the presence of two or more noise distributions, characterized by the typical Gaussian shapes, was appreciable.

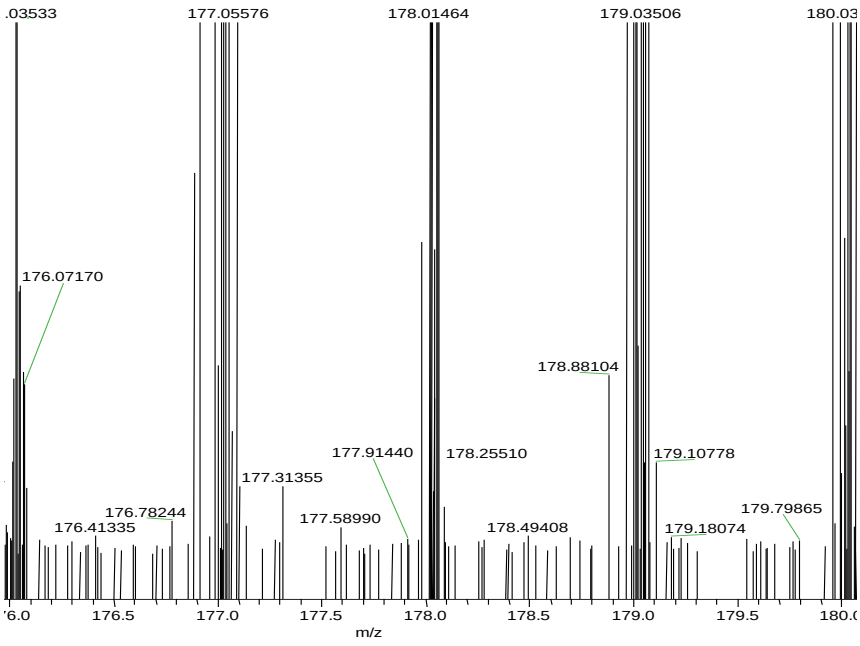

Figure 29. Mass spectrum detail showing the electronic noise within the *m/z* gaps.

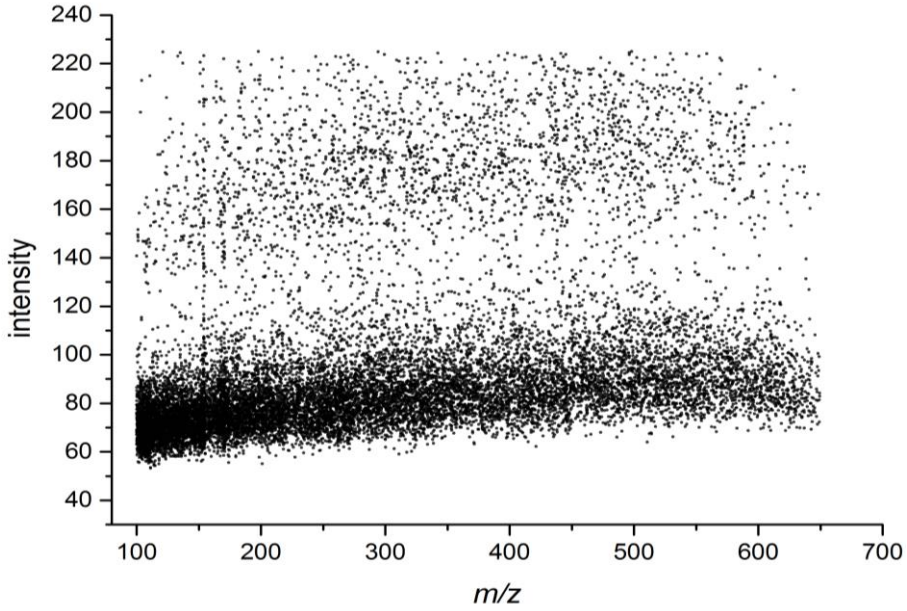

Figure 30. Noise signals distribution intensity vs. *m/z*.

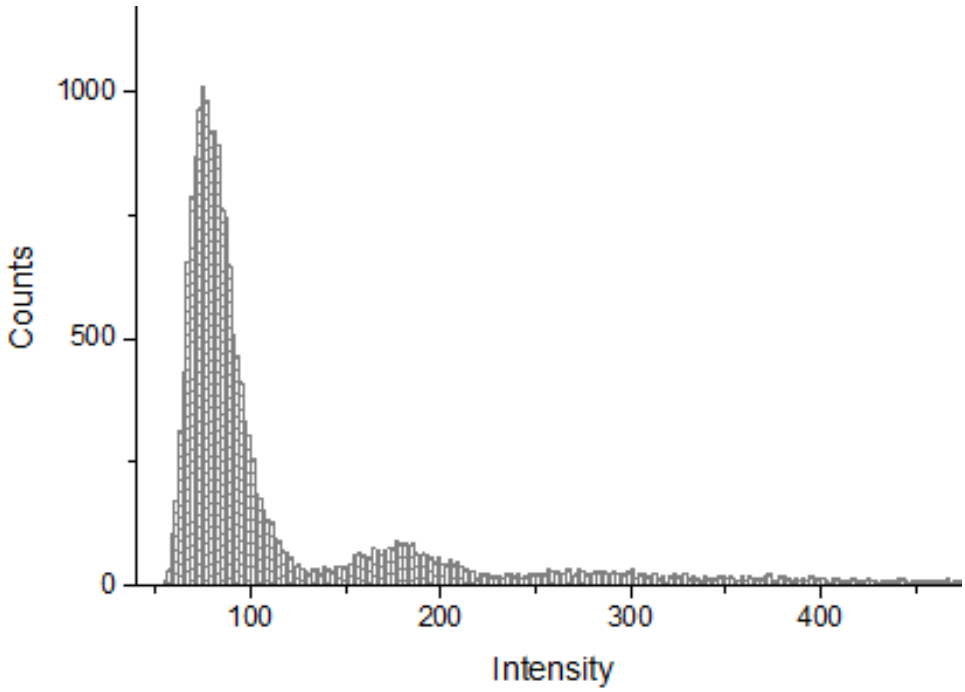

Figure 31. Histogram for the density distribution of the noise signals vs. the intensity.

The presence of more than one noise distribution is an artefact of the technologies employing the Fourier transform as Orbitrap™. Using different ion sources, the noise behaviour was similar, changing only in the position of the central values of the modes,

depending on the number of ions entering the analyser. In order to apply the standard formula for the signal to noise level (NL), the worst condition (the highest intensity of noise) has to be taken into account. The second noise distribution has a shape low and enlarged and its sampling resulted much harder to perform by software in comparison to the recognition of the main one. We have performed the calculation of NL simply multiplying the average of the main noise distribution by a factor of three. The level resulting is in every case greater than the second distribution (Figure 32).

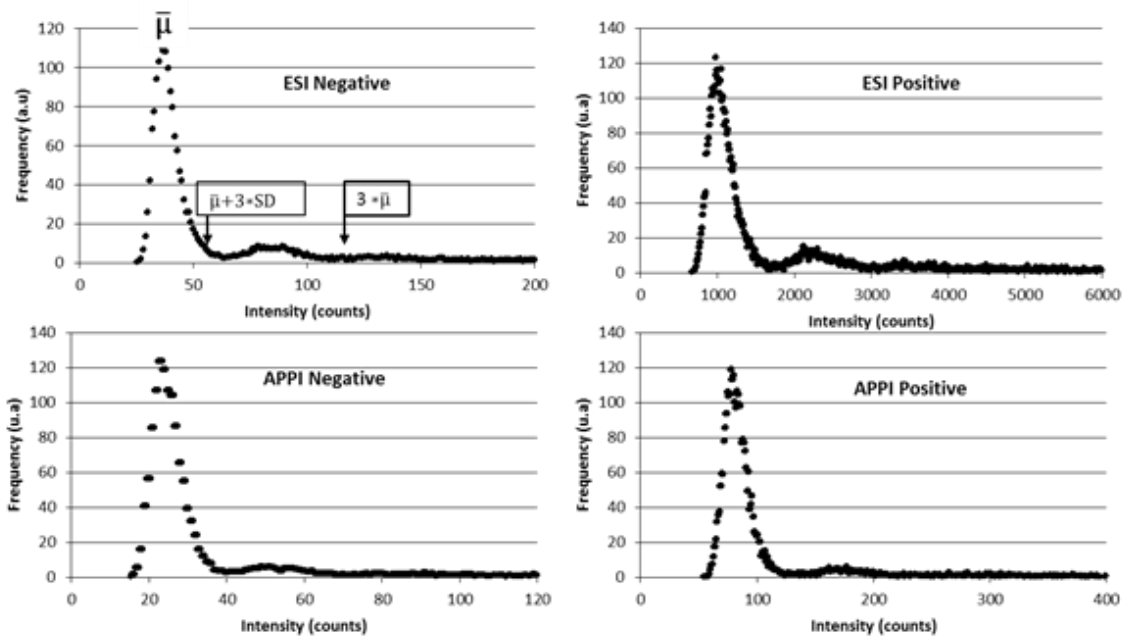

Figure 32. Noise signals vs. the intensity for different ion sources.

For each single replicate 27000 signals were commonly present. The signals having intensities below "LN= μ + 3*SD", were around 14000. At the contrary, the signals below "LN= 3*μ", were are 18000. Despite the great quantity of signals not considered using the major NL (~4000) when we perform the evaluation of the common ions throw the replicate the differences are really low and this is due to the characteristic of randomness typical of the noise signals.

In the Figure 33 are reported the van Krevelen plots of a single replicate related to a methanol extract of a winter sample of $PM_{2.5}$ are reported, showing a single replicate and the common ions obtained from three replicates with the two different NLs. Despite any loss of molecular formulas, the overall representation of the composition of the sample is not affected since the compounds present in the data analysis using "LN= μ +

3*SD" the standard deviation showed a random position in the plot. Thus, three times the mean value of the noise will used to avoid the second noise distribution and the possible presence of false-positive molecular formulas.

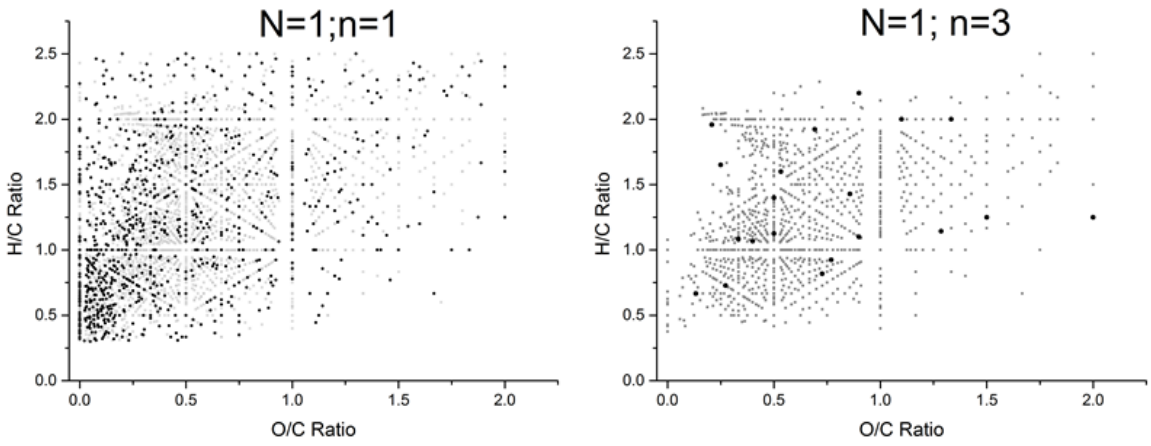

Figure 33. Van Krevelen plots after noise subtraction. Left) single sample and single replicate. Right) single sample and three replicates. Grey dots: NL= μ+3*SD; dark dots: μ+3*SD<NL<3*μ.

9.2.2.4 Blank subtraction

To avoid the presence of false-positive results, the blank subtraction process was moved at the beginning of the algorithm, after the noise removal. It was statistically carried out on the *m/z* values. In order to implement this step, two new parameters have been introduced to characterize the blank and sample mass spectra, calculated on the *m/z* values of known contaminants with MS signals covering the whole mass range of the acquisition.

- μ (mass accuracy): average of the mass errors associated to the known contaminants. This parameter, that can be also called mass bias, represents the mass accuracy of the measure and depends on the instrument calibration over the selected mass range.
- SD (mass precision): standard deviation of the contaminant mass errors. This parameter represents the precision of the instrumental measures and takes in to account of the variance function of the *m/z* value.

Blank and sample spectra could have different μ values. The algorithm first provides the correction of the *m/z* values subtracting the mass bias in accordance with the formula:

$$m/z_i^* = m/z_i - \frac{\mu_S \times m/z_i}{10^6}$$

The algorithm calculates the distance in ppm units between each corrected *m/z* values in the sample and each ones in the blank and then it performs a hypothesis t-test in accord with the formulas:

$$\Delta m/z = \frac{(m/z_B^* - m/z_S^*) \times 10^6}{\frac{(m/z_B^* + m/z_S^*)}{2}}$$

$$t_{s(0.05,\, df)} \geq \frac{\Delta m/z}{2SD_{comb}}$$

Where SD_{comb} is:

$$SD_{comb} = \sqrt{SD_B^2 + SD_S^2}$$

If the hypothesis t-test is true, which means signals of the sample and blank are statistically equivalent, the algorithm performs a test on the mutual intensities of the signals. If the ratio between the sample and blank signals is lower than the "sample to noise" value, here chosen equals to five, the algorithm deletes the *m/z* value.

9.2.2.5 Blank subtraction test comparison

A comparison between the original ("CodeESI+APPI_old") and modified (MassSpecProcessing v1.0 with the new instructions) algorithm was performed on a pair of real and related blank samples. This study was carried out in order to evaluate if the algorithm "MassSpecProcessing v1.0" is able to limit the presence of false-positive results arising from the possible scenarios described above by comparison the two list of filtered results. The sample chosen for this study was FP-1 analysed in positive APPI and negative nanoESI in the mass range of *m/z* 100-650. The constriction of mass error was adapted to be the same in the two different definitions: thus, in the elaboration with

"CodeESI+APPI_old", μ ± 2SD was used instead of the usual minimal value-0.5 ppm and maximum value+0.5 ppm. The used input parameters are reported in the Table 17.

Table 17. Input parameters for blank subtraction test.

		Positive APPI			Negative nanoESI		
CodeESI+ APPI_old		ppm upper	ppm lower	noise	ppm upper	ppm lower	noise
	Sample	-0.1425	-1.147	322	0.697	-0.161	250
	Blank	-0.3245	-1.293	384	0.333	-0.634	307
MassSpec Processing v1.0		Mass offset	SD	noise	Mass offset	SD	noise
	Sample	-0.6448	0.2511	322	0.268	0.215	250
	Blank	-0.8088	0.2421	384	-0.151	0.242	307

The main results are summarized in the Table 18:

Table 18. Time consumed for data analysis and detected molecular formulas using different algorithms of positive APPI sample and negative nanoESI sample. (N=1)

		CodeESI+APPI_old	MassSpec Processing v1.0
Positive APPI	time consumed	4'02"	1'10"
	output formulas	1020	947
	common formulas	946	
	unique formulas	74	1
Negative nanoESI	time consumed	20'18"	4'12"
	output formulas	3502	3420
	common formulas	3416	
	unique formulas	86	4

A first consideration has to be done concerning the elaboration time that was reduced to around a quarter using "MassSpecProcessing v1.0" instead "CodeESI+APPI_old". The post-run data analysis showed that "MassSpecProcessing v1.0" found always a lower number of results and that almost all the molecular formulas were also present in the list of results obtained with the previous algorithm. Conversely, the quite large number of molecular formulas found only by the "CodeESI+APPI_old" algorithm was significant. These results were carefully manually checked looking at the MS peaks of the real sample and the associated blank. As hypothesised, these false-positive resulted to be generated by the blank subtraction step, when carried out at the end of the elaboration process.

The presence of false positive results was mainly connected to the different assignation of molecular formulas on the slightly different *m/z* values. The Table 19 reports the formula assignations for a peak present in both sample and blank. The two *m/z* differed by 0.26 mDa corresponding to 1.30 ppm. The Xcalibur software, assigned a list of possible molecular formulas, and the formula "$C_7H_9O_2N_3S$" was present in both analysis but with different mass errors. "CodeESI+APPI_old" selected in the list of molecular formulas overcoming all the previous filters, those covering the range between minimal value-0.5 ppm and maximum value+0.5 ppm. Then, a probably incorrect assignment could be selected in the sample and not in the blank, and if the subtraction was not correctly performed a false-positive result was generated. Many other examples of this scenario were found in the results of this study.

Table 19. Assigned molecular formulas for positive APPI generating false positive results in the elaboration with the original algorithm.

	m/z	Intensity	Relative	Theoretical Mass	Delta (ppm)	RDB equiv.	Composition
Sample	199.0408	11032.2	0.07	199.04093	-0.65	1.5	$C_2H_9O_3N_634S$
				199.04100	**-0.99**	**5**	**$C_7H_9O_2N_3S$**
				199.04055	1.26	9.5	$C_{11}{}^{13}CH_8NS$
				199.04111	-1.57	4.5	$C_{10}H_{13}S^{34}S$
				199.04048	1.6	6	$C_6{}^{13}CH_8ON_4{}^{34}S$
				199.04037	2.15	1	$C_4{}^{13}CH_{10}O_8$
				199.04037	2.18	6.5	$C_3{}^{13}CH_4O_3N_7$
				199.04124	-2.21	0	$C_5{}^{13}CH_{14}O_3S_2$
				199.04002	3.95	0	$C_6{}^{13}CH_{16}S_2{}^{34}S$
				199.04165	-4.26	14	$C_{15}H_5N$
Blank	199.0405	10966.8	0.07	199.04055	-0.08	9.5	$C_{11}{}^{13}CH_8NS$
				199.04048	0.26	6	$C_6{}^{13}CH_8ON_4{}^{34}S$
				199.04037	0.82	1	$C_4{}^{13}CH_{10}O_8$
				199.04037	0.84	6.5	$C_3{}^{13}CH_4O_3N_7$
				199.04093	-1.99	1.5	$C_2H_9O_3N_6{}^{34}S$
				199.04100	**-2.32**	**5**	**$C_7H_9O_2N_3S$**
				199.04002	2.62	0	$C_6{}^{13}CH_{16}S_2{}^34S$
				199.04111	-2.9	4.5	$C_{10}H_{13}S^{34}S$
				199.0399	3.2	0.5	$C_3{}^{13}CH_{12}O_2N_3S_2$
				199.04124	-3.55	0	$C_5{}^{13}CH_{14}O_3S_2$

In the elaboration through "MassSpecProcessing v1.0" the blank subtraction was performed on the *m/z* values, since these two peaks were considered equal by the fact that their distance in ppm unit is 1.30, lower than the value of 1.43 ppm that was the

calculated confidence interval based on the precision and accuracy of the measures (SDcomb = 0.348).

Another scenario occurred during analysis was that two signals showed exactly the same *m/z* value and thus also the same list of associated molecular formulas. In this case, the false-positive result was provided by a different selection of formulas due to the different mass error range associated. In the example reported in the Table 20 "$C_{23}H_8ON_4$" was inside the range for the sample and not in the blank. Again, the subtraction was not performed correctly. Again, in the elaboration using "MassSpecProcessing v1.0" this result was not present because it was successfully subtracted considering the distance of the *m/z* values on the signals that in this particular case is 0.

Table 20. Assigned molecular formulas for negative nanoESI generating false positive results in the elaboration with the original algorithm.

	m/z	Intensity	Relative	Theoretical Mass	Delta (ppm)	RDB equiv.	Composition
Sample	356.06927	31601	0.2	**356.06926**	**0.03**	**22**	**$C_{23}H_8ON_4$**
				356.0693	-0.07	3.5	$C_9H_{18}O_6N_5S_2$
				356.06923	0.12	0	$C_4H_{18}O_7N_8S^{34}S$
				356.06932	-0.13	14	$C_{13}{}^{13}CH_9O_5N_7$
				356.06933	-0.15	8.5	$C_{14}{}^{13}CH_{15}O_{10}$
				356.06942	-0.4	3	$C_{12}H_{22}O_4N_2S_2{}^{34}S$
				356.06944	-0.46	13.5	$C_{16}{}^{13}CH_{13}O_3N_4{}^{34}S$
				356.06951	-0.65	17	$C_{21}{}^{13}C\ H_{13}\ O_2\ N\ S$
				356.06897	0.86	7.5	$C_{16}{}^{13}CH_{21}O_2S_2{}^{34}S$
				356.06885	1.18	8	$C_{13}{}^{13}CH_{17}O_4N_3S_2$
Blank	356.06927	31001.3	0.18	**356.06926**	**0.02**	**22**	**$C_{23}H_8ON_4$**
				356.0693	-0.09	3.5	$C_9H_{18}O_6N_5S_2$
				356.06923	0.1	0	$C_4H_{18}O_7N_8S^{34}S$
				356.06932	-0.15	14	$C_{13}{}^{13}CH_9O_5N_7$
				356.06933	-0.16	8.5	$C_{14}{}^{13}CH_{15}O_{10}$
				356.06942	-0.41	3	$C_{12}H_{22}O4N_2S_2{}^{34}S$
				356.06944	-0.47	13.5	$C_{16}{}^{13}CH_{13}O_3N_4{}^{34}S$
				356.06951	-0.66	17	$C_{21}{}^{13}CH_{13}O_2NS$
				356.06897	0.85	7.5	$C_{16}{}^{13}CH_{21}O_2S_2{}^{34}S$
				356.06885	1.17	8	$C_{13}{}^{13}CH_{17}O_4N_3S_2$

Definitely, we can assume that the previous algorithm was affected by the presence of false-positive results due to the presence in the sample and blank of peaks with: a) different *m/z* value and different mass error range of the measure; b) the same *m/z* value

and different mass error range of the measure. These issues were resolved in the "MassSpecProcessing v1.0" algorithm where the blank subtraction was carried out on the *m/z* values.

The presence of molecular formulas only found by "MassSpecProcessing v1.0", was connected to the low response of the ions in the sample and in the blank. An ion could be recognized in the sample and erased in the blank because below the noise level. Anyway, these molecular formulas were only a minor percentage of the result and did not affect the results.

9.2.3 μ, SD and noise quantifications: "Massdrift v.1.11" algorithm.

In order to implement the blank subtraction, two new metrics were introduced to characterize the mass spectra of the blank and the sample μ and SD.

The calculation of these two parameters were performed manually and it was extremely time consuming. In order to automatize the calculation an algorithm called "Massdrift v.1.11" was wrote and implemented in Mathematica.

Mass error offset (mass accuracy, μ) and standard deviation of mass errors (precision, SD) were automatically calculated using this algorithm, and it was based on known contaminants or substances likely to be present in the sample and previously confirmed via MS/MS experiment. The used signals and the characteristic fragmentation are reported in the Table 21.

This algorithm had as input a list of molecular formulas associated to each sample and blank, csv-file names and the respective csv-files. The lists of formulas were provided for both nanoESI and APPI; when a listed compound was not present in the spectrum, it did not affect the process. For each file listed, the algorithm searched in the related csv file the corresponding row for each individual formula, and rewrote in a table the associated error. If the code found more than one entry, it processed the peak with higher intensity. The algorithm calculated the average and SD of the found mass errors and applied the Grubbs test in order to evaluate the presence of outliers. In case of the presence of outliers, it removed them recalculating the average and standard deviation. The output table contained also the outliers marked with an asterisk in order to have all the information needed to manually check the results. Finally, the algorithm provided the number of molecular formulas used on the calculation as degree of freedom plus one.

Table 21. List of molecular formulas used in the mass accuracy and precision calculation and MS/MS characteristic fragments.

	exact *m/z*	ion formula	characteristic fragments
Positive	100.0757	$C_5H_{10}ON$	83.0490 (C_5H_7O) ; 72.0807 ($C_4H_{10}N$)
	149.0231	$C_8H_5O_3$	121.0282 ($C_7H_5O_2$)
	158.0963	$C_{11}H_{12}N$	143.0854 ($C_{11}H_{11}$)
	163.1227	$C_{10}H_{15}N_2$	132.0806 ($C_9H_{10}N$) ; 106.0650 (C_7H_8N) ; 84.0807 ($C_5H_{10}N$)
	107.0491	C_7H_7O	91.0542 (C_7H_7) ; 79.0542 (C_6H_7)
	123.0441	$C_7H_7O_2$	105.0335 (C_7H_5O) ; 95.0491 (C_6H_7O) ; 79.0542 (C_6H_7)
	125.0597	$C_7H_9O_2$	107.0491 (C_7H_7O) ; 97.0648 (C_6H_9O) ; 83.0491 (C_5H_7O) ; 79.0542 (C_6H_7)
	139.0753	$C_8H_{11}O_2$	107.0491 (C_7H_7O) ; 97.0648 (C_6H_9O) ; 79.0542 (C_6H_7)
	195.0876	$C_8H_{11}O_2N_4$	138.0661 ($C_6H_8ON_3$) ; 110.0713 ($C_5H_8N_3$)
	199.1117	$C_{14}H_{15}O$	181.1011 ($C_{14}H_{13}$) ; 171.1167 ($C_{13}H_{15}$) ; 166.0776 ($C_{13}H_{10}$) ; 155.0490 ($C_{11}H_7O$)
	227.2005	$C_{14}H_{27}O_2$	209.1899 ($C_{14}H_{25}O$) ; 199.0753 ($C_{13}H_{11}O_2$) ; 135.1168 ($C_{10}H_{15}$) ; 95.0855 (C_7H_{11})
	261.1849	$C_{17}H_{25}O_2$	219.1743 ($C_{15}H_{23}O$) ; 205.1223 ($C_{13}H_{17}O_2$) ; 177.1273 ($C_{12}H_{17}O$) ; 163.1116 ($C_{11}H_{15}O$)
	391.2841	$C_{24}H_{39}O_4$	297.1120 ($C_{18}H_{17}O4$) ; 279.15895 ($C_{16}H_{23}O_4$) ; 261.1484 ($C_{16}H_{21}O_3$) ; 149.0232 (C_8H5O_3)
negative	103.0042	$C_3H_3O_4$	75.0087 ($C_2H_3O_3$) ; 59.0138 ($C_2H_3O_2$)
	107.0503	C_7H_7O	79.0189 (C_5H_3O)
	123.0120	$C_3H_7O_3S$	108.9963 ($C_2H_5O_3S$) ; 90.9857 ($C_2H_3O_2S$) ; 76.9700 (CHO_2S)
	123.0452	$C_7H_7O_2$	121.0296 ($C_7H_5O_2$) ; 108.0217 ($C_6H_4O_2$) ; 95.0139 ($C_5H_3O_2$)
	139.0401	$C_7H_7O_3$	121.0296 ($C_7H_5O_2$) ; 111.0452 ($C_6H_7O_2$) ; 95.0503 (C_6H_7O)
	143.1080	$C_8H_{15}O_2$	125.0972 ($C_8H_{13}O$) ; 99.11792 (C_7H_{15})
	157.1234	$C_9H_{17}O_2$	139.0765 ($C_8H_{11}O2$) ; 127.1128 ($C_8H_{15}O$) ; 97.0660 (C_6H_9O)
	171.1392	$C_{10}H_{19}O_2$	153.1284 ($C_{10}H_{17}O$) ; 127.1128 ($C_8H_{15}O$)
	199.1704	$C_{12}H_{23}O_2$	181.1234 ($C_{11}H_{17}O_2$) ; 155.1442 ($C_{10}H_{19}O$) ; 137.0973 ($C_9H_{13}O$)
	227.2017	$C_{14}H_{27}O_2$	209.1548 ($C_{13}H_{21}O_2$) ; 183.1755 ($C_{12}H_{23}O$) ; 165.1285 ($C_{11}H_{17}O$)
	255.2330	$C_{16}H_{31}O_2$	237.2225 ($C_{16}H_{29}O$) ; 211.2069 ($C_{14}H_{27}O$) ; 193.1599 ($C_{13}H_{21}O$)
	283.2279	$C_{17}H_{31}O_3$	265.1807 ($C_{16}H_{25}O_3$) ; 239.2378 ($C_{16}H_{31}O$) ; 221.1909 ($C_{15}H_{25}O$)
	311.2956	$C_{20}H_{39}O_2$	279.2327 ($C_{18}H_{31}O_2$) ; 267.2691 ($C_{18}H_{35}O$) ; 249.2221 ($C_{17}H_{29}O$)
	339.3269	$C_{22}H_{43}O_2$	295.3002 ($C_{20}H_{39}O$) ; 277.2533 ($C_{19}H_{33}O$)

As the Table 22 and the Figure 34 show for an acquisition in ESI negative, the variability of the single mass error value through the 40 measures was always lower than the variability of the mass error through the mass range. In other words, the implemented approach takes into account the mass drift that we have along the increased *m/z* values. On the other hand, it is well known the characteristic of the

Orbitrap™ technology to have a good mass stability represented by the low standard deviation on the single mass throughout the all acquisition range.

Table 22. Mass errors and standard deviations associated to the selected molecular formulas of the single scans and comparison of the interval calculation.

mass variability thought the single acquisitions			
Formula	Theoric *m/z* value	mass error average on 40 measures	Standard deviation on 40 measures
$C_3 H_3 O_4$	103.00368	**0.47**	**0.10**
$C_6 H_4 O_3$	124.01659	0.36	0.09
$C_7 H_5 O_3$	137.02442	0.54	0.08
$C_8 H_{15} O_2$	143.10775	0.28	0.12
$C_6 H_4 O_4 N$	154.01458	0.25	0.11
$C_7 H_6 O_4 N$	168.03023	0.30	0.13
$C_8 H_8 O_4 N$	182.04588	0.21	0.13
$C_9 H_{10} O_4 N$	196.06153	0.31	0.11
$C_{12} H_{23} O_2$	199.17035	-0.04	0.22
$C_{14} H_{27} O_2$	227.20165	0.12	0.19
$C_{10} H_{15} O_6$	231.08741	**-0.10**	**0.14**
$C_{16} H_{31} O_2$	255.23295	0.08	0.17
$C_{18} H_{35} O_2$	283.26425	-0.07	0.19
$C_{17} H_{27} O_3 S$	311.16864	0.25	0.09
maximum value + standard deviation		0.64	
minimum value - standard deviation		-0.25	
mass error variability on the averaged spectrum			
mean value on average mass + 2 standard deviation		0.64	
mean value on average mass - 2 standard deviation		-0.25	

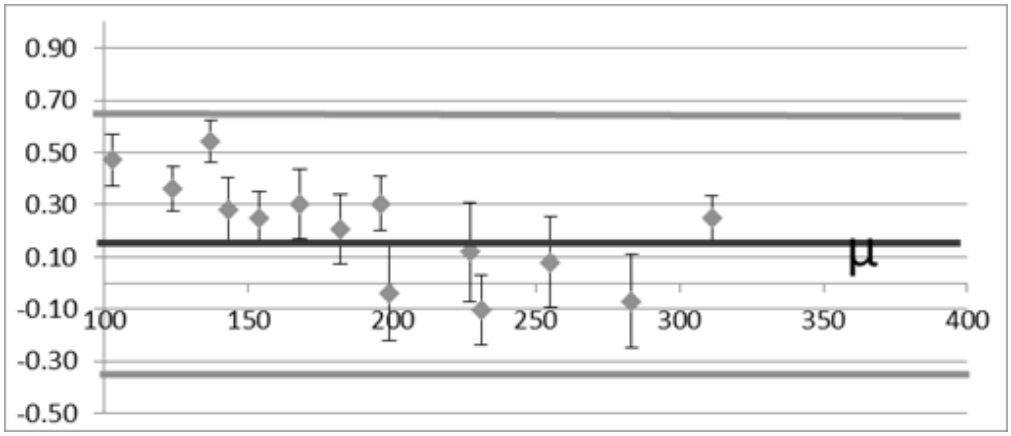

Figure 34. Mass error distribution vs *m/z* range.

μ value, it represents an offset of the mass calibration of the instrumentation, and the correction by this value has to be intended as a post-run recalibration. In general, experimental evidences have showed a negative correlation between the mass and associated error. This is properly called mass drift and will be take into account in the future advantages of the algorithm by fitting the mass errors with different equations and applying consequently a better mass correction.

The noise was quantified by the same algorithm as three times the average of the central value of the main normal distribution of counted signals per unit of intensity. The algorithm simply divided the mass spectra in bins of intensity and counts the ions present in them. The bins width is tuneable based on the overall intensity of the noise and the algorithm limited the noise calculation on the signals below an intensity level in order to save time. The algorithm plotted the found values versus the intensity, and fitted the data using the normal distribution equation. In the output table the mean value of the distribution and the associated standard deviation were generated. Over more than 450 spectra elaborations, no failures were recorded.

9.2.4 Merging of the mass scan range and determination of common ions.

Each single sample was recorded in two different *m/z* ranges (*m/z* 100-650 and *m/z* 150-900) in order to increase the ion transmission (function of *m/z* range) and the measure was replicate three times. After the elaboration with the algorithm "MassSpecProcessing V1.0", data were exported as .csv files. These lists contained single-*m/z* value formula associations. The successive step was merging results from the

different MS scan ranges (*m/z* 100-650 and 150-900). Then, in order to limit the presence of false-positive results, only formulas experimented in all the three replicates were selected. This simple precaution has a significant positive effect on the reliability of results. The determination of the common ions achieved the elimination of all non-repeatable formula association, which could arise by:

a) The non-presence of a signal due to the intensity cut-off.
b) The noisy nature of signals.
c) The non-reproducible formula assignation.

The point a) is a direct consequence of two main factors. First, the response of a compound can vary even a lot from a measure to one another because is sensitive to ionization suppression effects due to the matrix and in the case a compound is just at low concentration that suppression effects could break down the signals with consequent no-presence in the spectra. Second, we had introduced in the algorithm a preliminary intensity cut-off level. This means that a signal has to be above that level in all the three measure; the matrix effects act in same way as before.

For the issue related to the point b, we can have the presence of two type of noise signals. Electronic noise is independent by the mass and its intensity distribution is a Gaussian; the noise cut off introduced deleted the majority of the noise signals but not all. High intensity noise signals were present also at the bases of the dominant peaks in what is called "shouldering". This artefact has an electronic-mathematic origin and it is a consequence of the use of centroid spectra. In fact, in the FT instrumentation, the limitation in the time acquisition creates an enlargement of the base of profile peaks that is reconstructed by centroids. Both Electronic and shouldering peaks are characterized by *m/z* values highly variable and the simultaneous presence of the same signal in the three replicates is improbable.

The last point is the main cause of presence of false-positive results. Conceptually, when a molecular formula is associated to a *m/z* value, the probability that the formula is true drop if the associated error is far from the μ value of the measure (mass drift). Consequently, in different measures it is highly probable that the same association is not present because it was excluded by the standard deviation range. The common ion determination was a powerful tool to limit the presence of these kind of results.

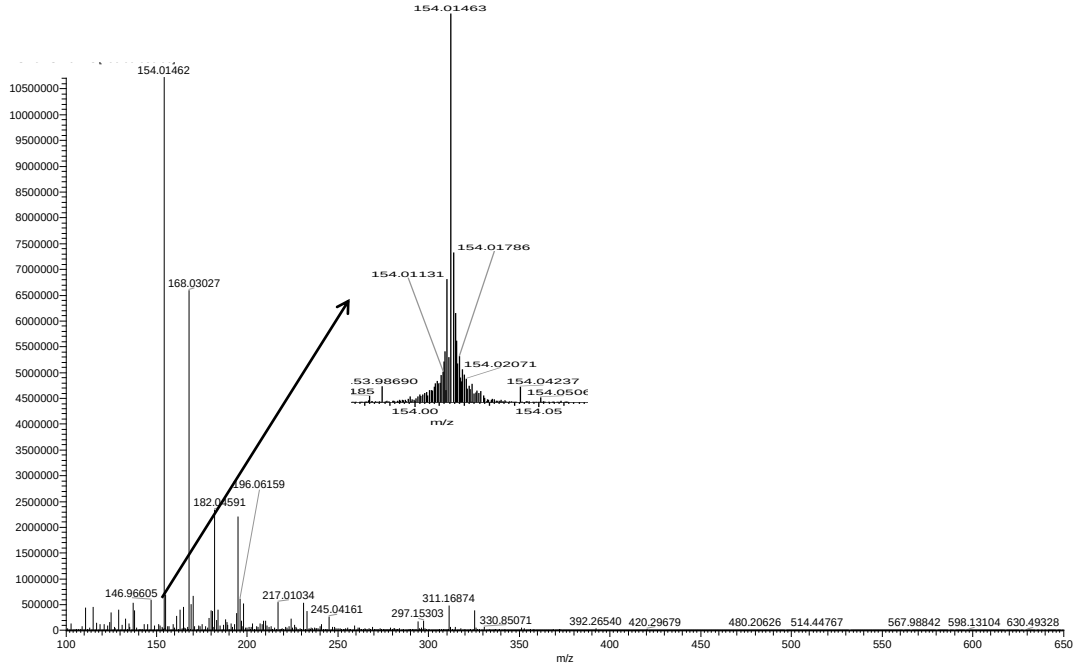

Figure 35. Classic shouldering of dominant peak.

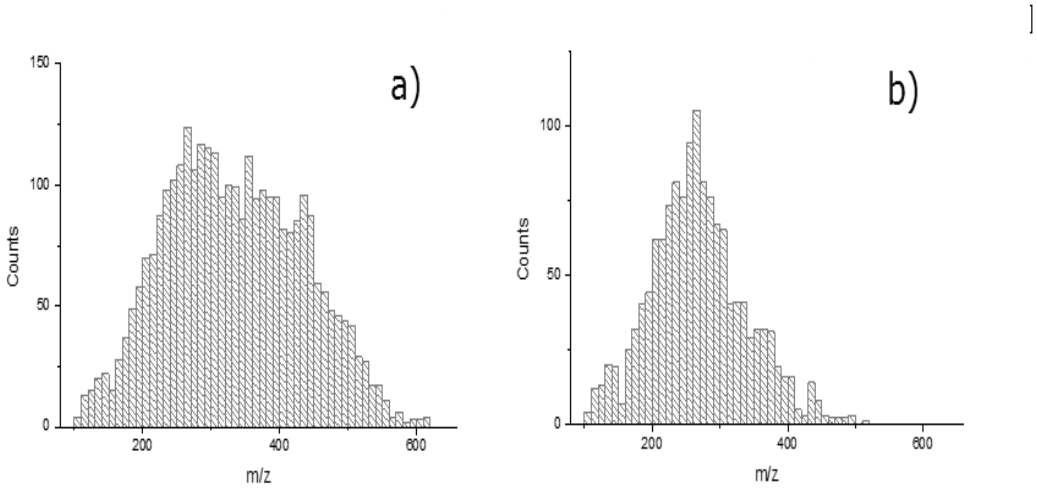

Figure 36. Distributions of results over the *m/z* range 100-650: a) ions in a single measure; b) common ions in the three replicates.

9.3 Application of the protocol on real samples

9.3.1 Choice of the samples and analysis

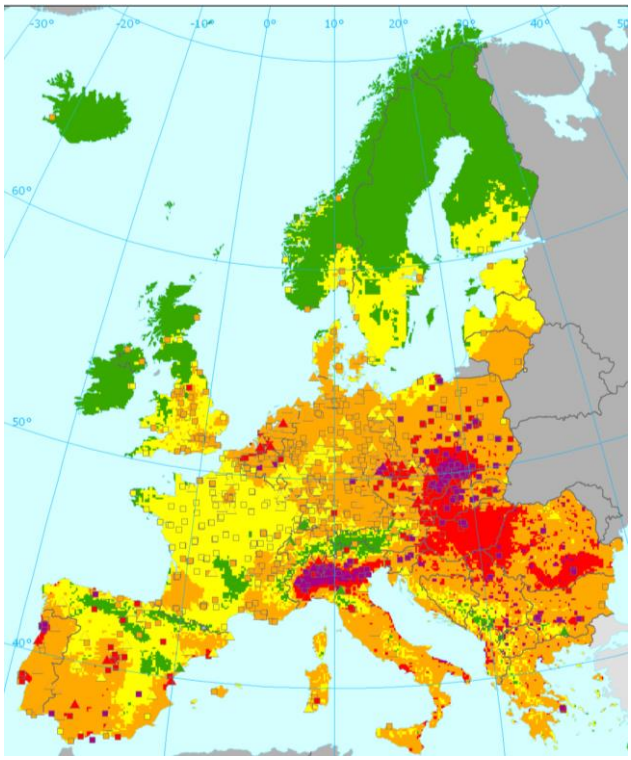

Figure 37. PM concentration level in Europe.

Po valley represents a hot spot in Europe for the concern of particulate matter pollutions (Figure 37) [107, 108]. In fact, in this region, the main PM concentration is high with exceeding of the legal threshold value often occurring during the winter season. The choice of that area ensured the presence in the sample of a high concentration of $PM_{2.5}$ and consequently a good concentration of organic compounds in the extract.

Ten real samples, five from winter and five from summer monitoring, and two blanks were analysed; two aliquots of each sample were extracted for the determination by using APPI and nanoESI sources. Considering that filter aliquots were analysed with two polarities, two mass ranges and in triplicates, a total 288 spectra were collected. Despite of this extremely high number of spectra, the data analysis performed with the new protocol required a total of only 18 hours.

In general, the large majority of the peaks in the spectra was below *m/z* 350 with only few peaks between *m/z* 350 and 450, and almost no peaks at *m/z*>450. These results were in contrast with experiments attempted on aerosols generated in a smoke chamber [109] or collected in far-from civilization areas as Amazonia [110], where high molecular weight compounds were reported. In positive nanoESI acquisition, $[M+Na]^+$ adducts represented about 13% of the total assignations, and in both positive and negative APPI acquisition molecular ions were one fifth of the total signals.

9.3.2 Number of detected compounds

The first parameter highlighted was the number of molecular formulas identified by the protocol in the real sample extracts. The results for the samples, winter and summer, are reported in the Table 23 as average of detected formulas on the five samples.

Table 23. Averaged number of detected molecular formulas (N=5).

Winter samples				Summer samples			
NanoESI		APPI		NanoESI		APPI	
Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
2390	1826	1691	1019	1985	1545	187	123

In general, nanoESI was able to detect a larger number of formulas than APPI both in winter samples and summer samples, even if with a significant difference in magnitude. However, while in the winter sample the results amount was comparable between the two sources, in the summer sample the response of APPI was poor.

The significant lower number of compounds evidenced by the APPI analysis of the summer sample could be explained by the absence of compounds that can be photo-ionized, and this fact could arise from:

- The non-emission of this class of compounds.
- The presence of degradation reactions.
- The non-condensation or absorption of the compounds on the aerosol.

We considered more likely the last two hypothesis, since it is known that the larger amount of light in the summer season could promote photo-degradation phenomena of pollutants in the atmosphere and, at the same time, the higher temperature speeds up the reactions and modifies the partition constants of the substances between the air and the particles.

Last consideration about the number of formulas detected is about the general lower response in positive polarity with respect of the negative one. This could be correlated to the higher noise level of the spectra recorded in positive mode.

.

9.3.3 Chemical composition

The large datasets of molecular formulas obtained were used characterize the $PM_{2.5}$ in terms of composition classes.

Table 24. Elemental composition of the detected molecular formulas.

	Winter samples			
Composition class	NanoESI POS	NanoESI NEG	APPI POS	APPI NEG
Tot	1058	958	466	911
CH	5	0	14	0
CHO	357	346	282	421
CHON	568	506	153	418
CHN	128	1	13	68
CHSON	0	25	2	4
CHSO	0	80	2	0
	Summer samples			
Composition class	Nano ESI POS	Nano ESI NEG	APPI POS	APPI NEG
Tot	713	1091	75	100
CH	4	0	0	0
CHO	282	275	51	28
CHN	30	1	0	0
CHON	386	520	18	52
CHONS	11	106	5	15
CHOS	0	189	1	5

The Table 24 refers on the number of the common ions between the five samples in the two different seasons. The same results can be displayed in pie charts (Figure 38) that showing graphically the composition making easier and immediate the data comprehension.

APPI results for summer sample are not reported because of the low number of detected molecular formulas.

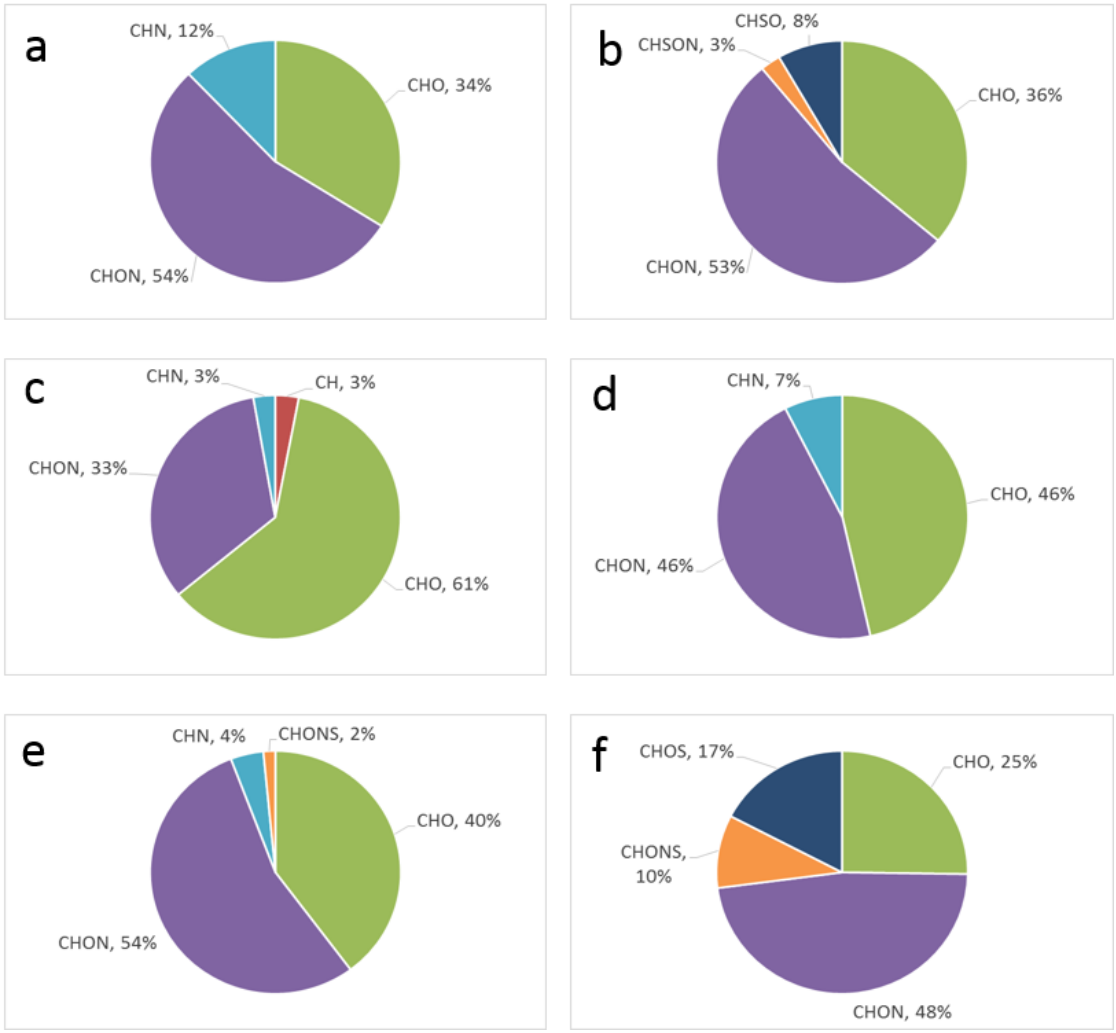

Figure 38. Pie chart representing the chemical class composition of winter sample: a) positive nanoESI, b) negative nanoESI, c) positive APPI, d) negative APPI and summer sample: e) positive nanoESI, f) negative nanoESI.

Several sub-groups of molecular formulas were highlighted, and results of the MS analysis performed with different ion sources can be here summarized:

- Positive nanoESI efficiently underlined CHN compounds that represented 12% of the total $PM_{2.5}$ composition. The contribution of this class strongly decreased in the summer samples.
- Negative nanoESI was sensitive for compounds containing sulphur. The percentage of the sum of the sub-groups CHSO and CHNSO increased from 11% in the winter $PM_{2.5}$ extract to 27% of the summer sample.
- In general, APPI did not show a significant presence compounds different from the CHO and CHNO sub-classes. Despite this, it was notable that 3% of APPI in positive acquisition mode was referred to the CH sub-class, and CHN compounds

were detected in negative polarity (7%) instead of to the positive mode of the nanoESI.

A general statement of the atmospheric science is that in winter the chemistry of the atmosphere is mainly correlated to the chemistry of NO_x, while in the summer to the SO_2 chemistry. As general inference, from the study of elemental sub-groups composition of our seasonal sample analysed in negative nanoESI, we have confirmed for the urban environment the dominant presence of compound containing sulphur (CHSO and CHSNO), whilst in the summer sample a greater presence of CHNO compounds was experienced. However, as remarkable evidence, positive nanoESI did not show any difference in the percentage of CHNO compounds between winter and summer sample. This indicates a chemistry of the PM based mainly on non-oxidized nitrogen.

9.3.4 Van Krevelen plot analysis

Van Krevelen plots for winter sample (Figure 39 a, b, c, d) and summer sample (Figure 39 e, f) are displayed in 3D version using as third dimension a scale of colours based on the DBE value associated to the formulas. Some highlights arising from the evaluation of van Krevelen diagrams are here reported:

- While negative polarity covers a wide range on O/C ratio (until 2), positive analysis produces formulas with O/C ratio relatively low (below 1).
- APPI and nanoESI plots exhibit the same morphology concerning the polarity of acquisition for the winter sample.
- Negative nanoESI for winter and summer samples showed very few formulas with O/C ratio equals to 0. This evidence was in accordance with the ionization rule for electrospray source: presence of oxygen on the structure is necessary in the ionization for the stabilization of the negative charges.
- Positive nanoESI exhibited a large number of formulas without oxygen, and mainly due to the presence of the CHN fraction.
- APPI positive and negative plots for winter sample showed the presence of formulas with O/C equals to 0. The photoionization does not need oxygen in the structure to form ions. Conversely, unsaturations and π-delocalizations are needed. The plots exhibit a wide number of formulas characterized by a high value of DBE.

- While the covered area of the plots of winter and summer samples are the same in the nanoESI, the DBE values associated to the formulas are drastically different. In fact, winter sample showed the presence of high DBE compounds.

In positive nanoESI was evident that the area characterized by high H/C is richer of formulas in the summer sample than in the winter one.

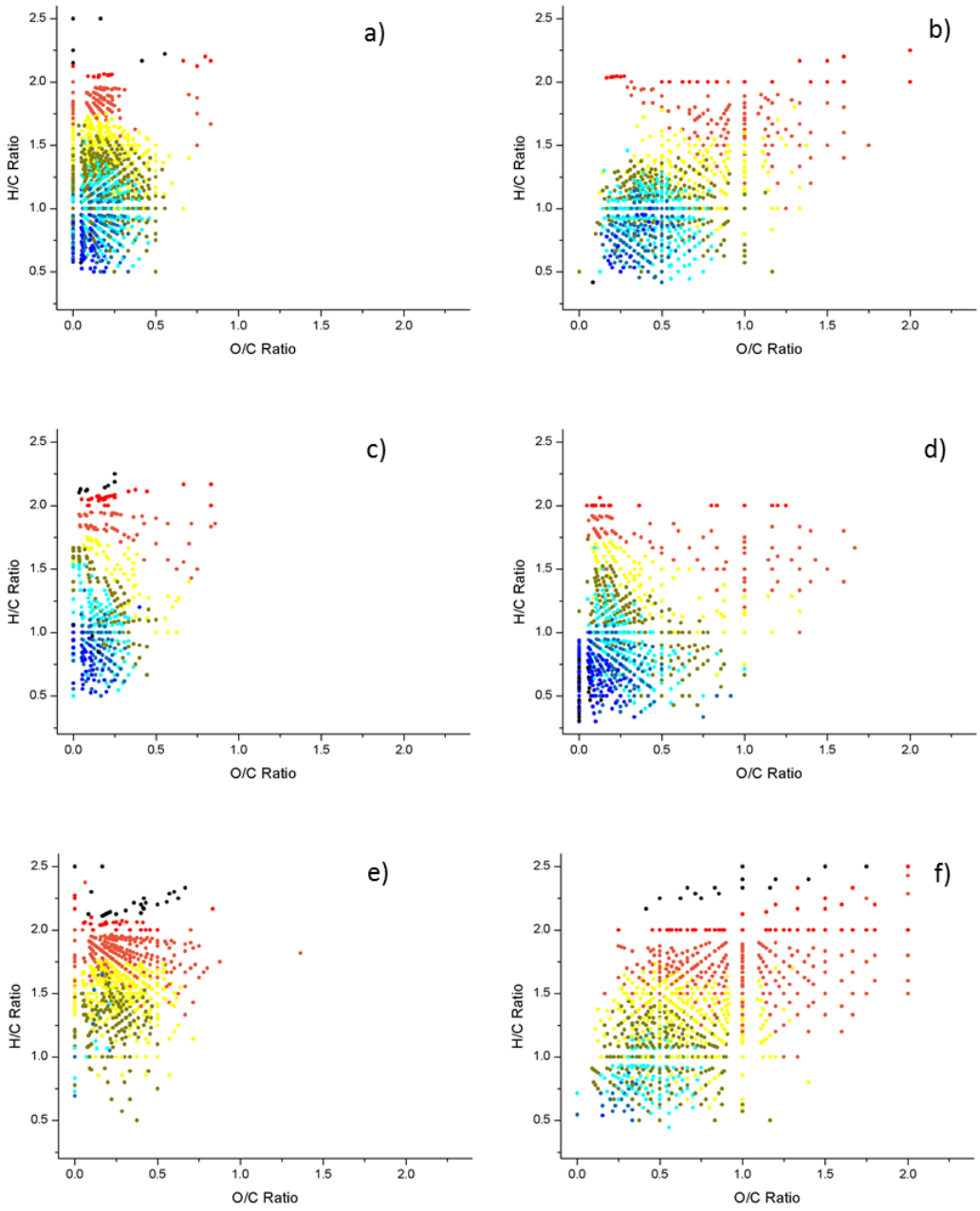

Figure 39. Van Krevelen plots of winter sample: a) positive nanoESI, b) negative nanoESI, c) positive APPI, d) negative APPI and summer sample: e) positive nanoESI, f) negative nanoESI.

9.3.4.1 Density van Krevelen plot

In van Krevelen plots, the information is provided by the position in the plane of the formulas with the consequent necessity of evaluation of specific regions and the presence of characteristic patterns. In order to simplify the evaluation of these diagrams we have introduced the density van Krevelen plot reported in the Figure 40 for winter (a, b, c, d) and summer samples (e, f). The densities were calculated implementing in Origin™ the kernel data analysis, and setting bins of 0.2 in both the axes.

In this version of the plots, the information desumable is the morphology of the formula density and the position of the maxima. The latter is similar to the averaged values of the descriptors reported in the Table 25.

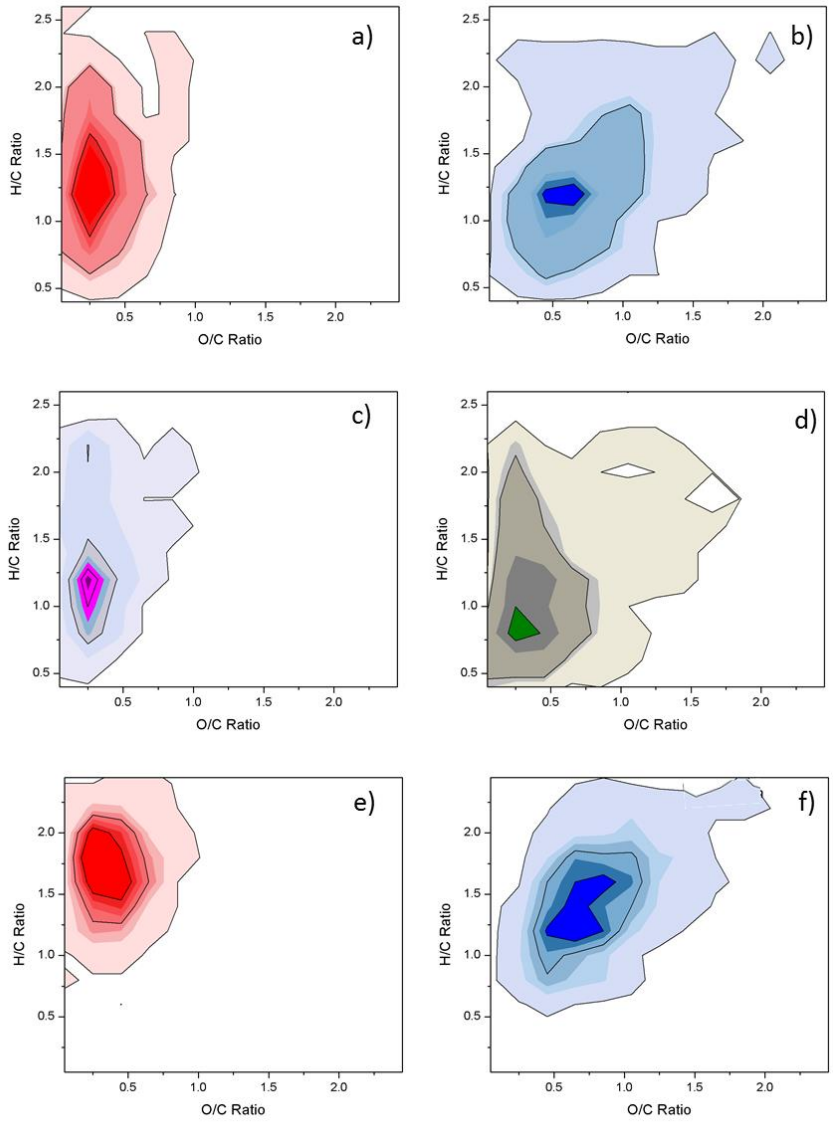

Figure 40. Density van Krevelen plot for winter sample: a) positive nanoESI, b) negative nanoESI, c) positive APPI, d) negative APPI and summer sample: e) positive nanoESI, f) negative nanoESI.

Table 25. Mean values of H/C, O/C and DBE of the data for winter and summer sample collected using different ion sources and polarity.

Winter												
ESI NEG			ESI POS			APPI NEG			APPI POS			
O/C	H/C	DBE	O/C	H/C	DBE	O/C	H/C	DBE	O/C	H/C	DBE	
0.60	1.13	6.44	0.19	1.21	7.07	0.32	1.00	7.69	0.20	1.26	7.01	
Summer												
ESI NEG			ESI POS			APPI NEG			APPI POS			
O/C	H/C	DBE	O/C	H/C	DBE	O/C	H/C	DBE	O/C	H/C	DBE	
0.73	1.35	4.63	0.28	1.58	4.13	0.32	1.44	5.35	0.20	1.55	4.68	

A first consideration has to be mentioned about APPI plots of winter sample: positive and negative ionization polarities seem to produce similar representations, and did not give any additional information with respect of those obtained with the positive nanoESI.

The density plots of nanoESI results, in combination with the mean values, easily show some interesting differences between the two seasons: in summer sample, the chemical profiles are characterized by greater values of O/C and H/C ratios. In the negative polarity, formulas containing oxygen responded better and consequently the increase in O/C was well displayed. Conversely, the positive nanoESI has proved to be sensitive to the H/C ratio, with a notable ability to show its increment..

9.3.5 Oxidation state of the carbon

As it can be seen for van Krevelen plot, the oxidation state of the carbon plot is particular useful for the implementation of the kernel density in order to have an understandable representation. The plots in the Figure 41 display both scatter and density representations and on the x-axes, the number of carbons has been chosen.

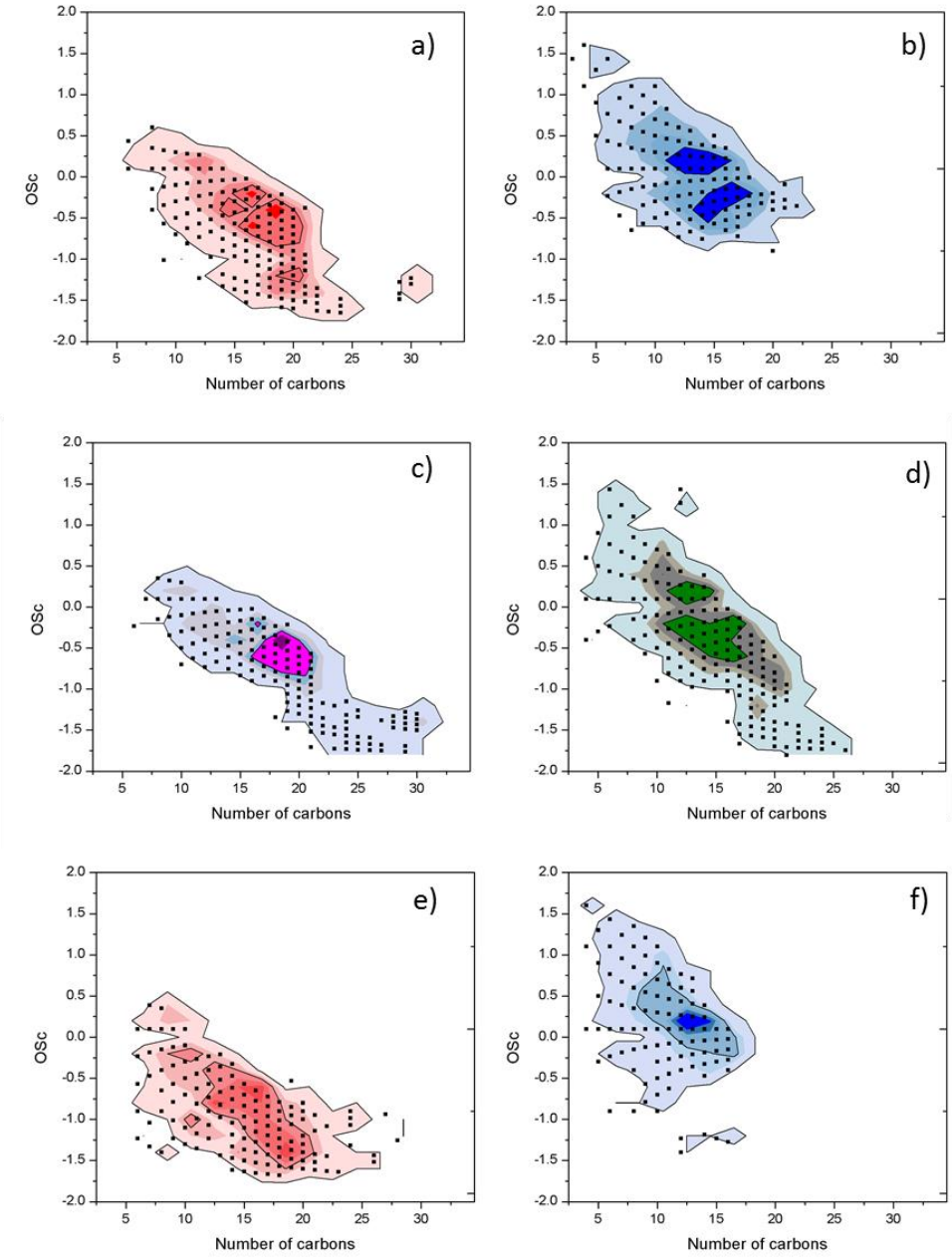

Figure 41. Oxidation state of the carbon vs. number of the carbons density plot for winter sample: a) positive nanoESI, b) negative nanoESI, c) positive APPI, d) negative APPI and summer sample: e) positive nanoESI, f) negative nanoESI.

In general, positive polarity for both nanoESI and APPI in the two seasonal sample, covered an area in the plot characterized by high values of OSc and low number on

carbon, while positive polarity detected higher molecular weight low oxidized structures.

The study of the oxidation state of the carbon is particular useful for the comprehension of reactions involved in the formation and transformation of the aerosols. For studying the different processes involved during the two seasons, we have plotted (Figure 42) all the formulas obtained by nanoESI in positive and negative polarities. While the winter distribution appears much more compact, the summer composition is spread on the diagram plane. This was due to the presence of reactions of fragmentation and functionalization increasing the oxidation of the structures and simultaneously decreasing the molecular weight. A second contribute was imputed to the reactions of oligomerizations, which drive the formation of high molecular weight structures, characterized by a slightly lower OSc.

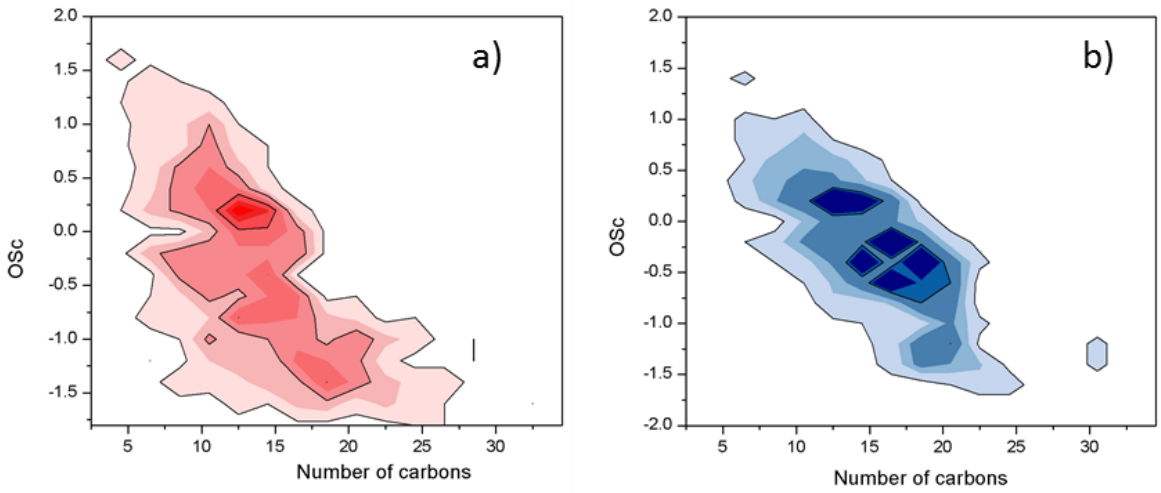

Figure 42. Oxidation state of the carbon vs. number of the carbons density plot for: a) winter sample positive and negative nanoESI, b) summer sample positive and negative nanoESI.

If we consider the single contributions of each polarity in nanoESI, reported in the graft Figure 43 and Figure 44, it is possible to observe that:

- Negative nanoESI is able to characterize reactions of fragmentation and functionalization.
- Positive nanoESI is more sensitive to the presence of oligomerization reactions.
- In winter aerosols the emission contributions are dominant
- Summer aerosol composition is strongly affected by the aging reactions.

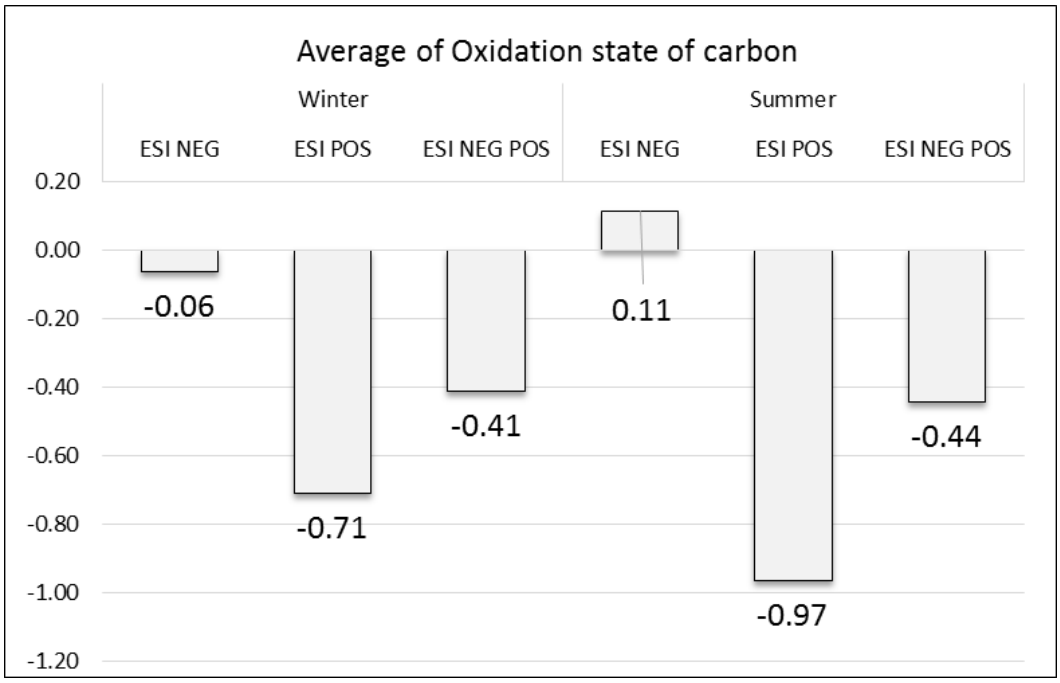

Figure 43. Mean values of OSc for the winter and summer samples analysed by nanoESI.

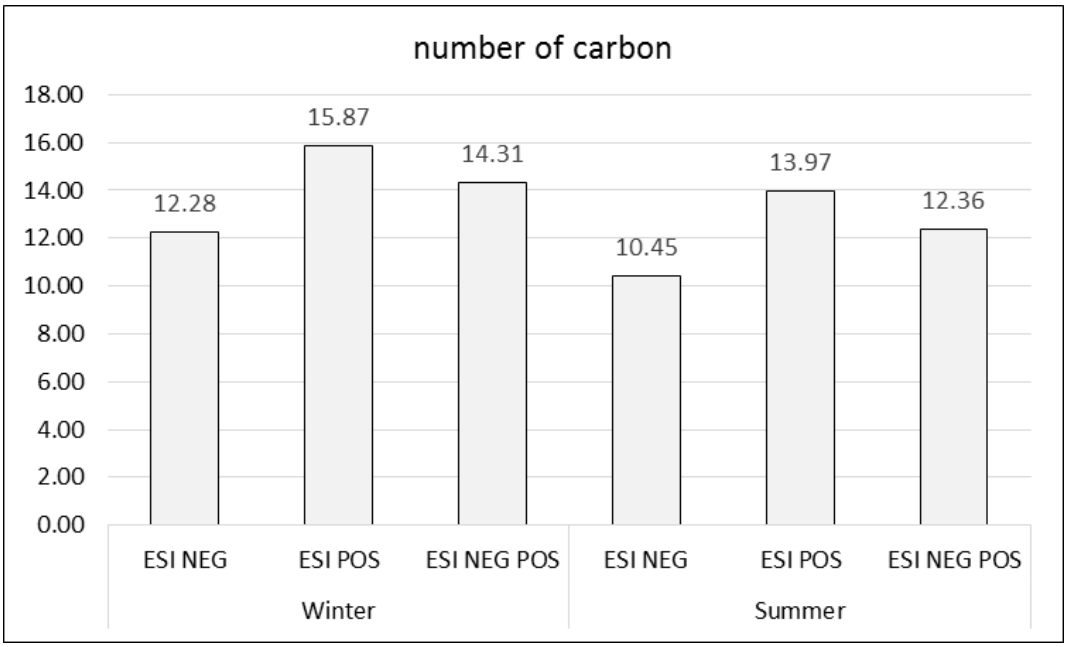

Figure 44. Mean values of number of carbons for the winter and summer samples analysed by nanoESI.

9.3.6 Kendrick mass defects analysis

Homologue series in Kendrick mass defect plots form horizontal lines on the plane (Figure 45). By evaluating the plots and the data it is possible to obtain a tentative chemical classification of the molecular formulas, thus increasing the identification confidence. The Table 26 reports the lengthiest homologue series for the winter sample analysed by nanoESI in positive and negative mode.

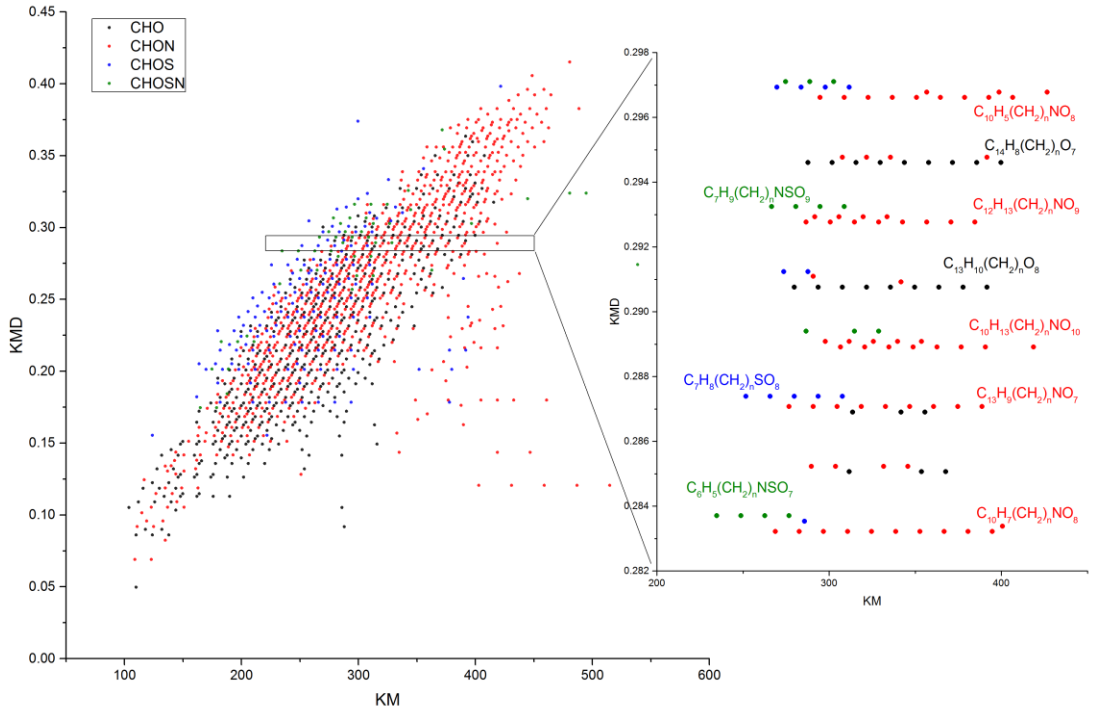

Figure 45. Example of Kendrick mass defect plot for winter sample analysed by positive nanoESI

As we can see in the Table 26, for winter sample analysed in nanoESI, the chemical classes present with lengthiest homologue series in negative polarity, are mainly sulphur-containing compound. Organosulfates (SO_4) and sulphuric acid derivatives (SO_3) are the most likely members of these sub-groups. Other detected long series were formed by CHNO compounds as nitro-derivatives and CHO as carboxylic and dicarboxylic acids. Lengthiest series of positive nanoESI were mainly represented by CHNO and CHO compounds.

Table 26. Homologue series for the winter sample analysed by nanoESI in positive and negative mode

	First member of the homologue series	KMD	Carbon range of the series	Length of the series
Negative nanoESI	$C_3H_6SO_6$	0.201299	C_3-C_{26}	22
	$C_4H_6SO_6$	0.214698	C_4-C_{25}	21
	$C_6H_{11}NO_7$	0.179877	C_6-C_{26}	18
	$C_5H_7NO_8$	0.216221	C_5-C_{22}	18
	$C_4H_8SO_7$	0.224244	C_4-C_{22}	18
	$C_4H_6SO_7$	0.237643	C_4-C_{24}	18
	$C_5H_7NO_7$	0.193276	C_5-C_{21}	17
	$C_5H_{10}N_2O_9$	0.231629	C_5-C_{22}	16
	$C_4H_6SO_8$	0.260588	C_4-C_{25}	15
	$C_5H_5NO_7$	0.206676	C_5-C_{20}	14
	$C_6H_9NO_9$	0.239167	C_6-C_{20}	14
	$C_7H_{12}N_2O_9$	0.245029	C_7-C_{20}	14
	$C_4H_4SO_7$	0.251042	C_4-C_{20}	14
	$C_7H_{10}N_2O_9$	0.258428	C_7-C_{20}	14
	$C_5H_6SO_8$	0.273987	C_5-C_{20}	14
	$C_2H_4SO_4$	0.155409	C_2-C_{27}	13
	$C_5H_{10}SO_5$	0.178354	C_5-C_{26}	13
	$C_6H_7NO_8$	0.229621	C_6-C_{18}	13
	$C_8H_9NO_9$	0.265965	C_8-C_{21}	13
	$C_8H_4O_4$	0.172176	C_8-C_{19}	12
	$C_5H_4O_7$	0.200813	C_5-C_{18}	12
Positive nanoESI	$C_8H_9NO_4$	0.15124	C_8-C_{32}	20
	$C_8H_7NO_3$	0.141694	C_8-C_{28}	17
	$C_8H_7NO_4$	0.164639	C_8-C_{29}	16
	$C_9H_9NO_5$	0.187584	C_9-C_{29}	16
	$C_6H_8O_3$	0.095634	C_6-C_{26}	15
	$C_6H_5NO_2$	0.10535	C_6-C_{29}	15
	$C_8H_4O_2$	0.126286	C_8-C_{28}	15
	$C_9H_{11}NO_5$	0.174185	C_9-C_{28}	15
	$C_{11}H_9NO_5$	0.214383	C_{11}-C_{25}	15
	$C_6H_5NO_3$	0.128295	C_6-C_{21}	14
	$C_{10}H_6O_2$	0.139685	C_{10}-C_{30}	14
	$C_{10}H_6O_3$	0.16263	C_{10}-C_{30}	14
	$C_{14}H_{24}O_2$	0.072689	C_{14}-C_{27}	13
	$C_9H_7NO_2$	0.132148	C_9-C_{30}	13
	$C_9H_6O_3$	0.149231	C_9-H_{21}	13
	$C_6H_{10}N_2$	0.038523	C_6-C_{32}	13
	$C_8H_{10}N_2$	0.065322	C_8-C_{16}	12

A representation of the chemical complexity of the composition could be evaluated by the compounds per series ratio reported in the Table 27. We can see for example, that

this descriptor decrease from 6.94 to 3.69 in negative nanoESI from the winter to the summer sample and, at the contrary, in positive nanoESI we have an increment in the value from 3.38 to 5.35. These results corroborates again the results obtained in the analysis of OSc plots: reaction of oligomerization, fragmentation and functionalization affected the composition in summer $PM_{2.5}$,while in winter $PM_{2.5}$ primarily emitted compound are the major contribution.

Table 27. Formulas, series and formula to series ratio for winter and summer sample analysed by nanoESI.

			Number of formulas	Number of series	Formulas per series
nanoESI	Winter	Negative	2390	344	6.94
		Positive	1826	540	3.38
	Summer	Negative	1985	537	3.69
		Positive	1545	288	5.35

9.4 Conclusions

In order to assign the most probable molecular formulas to the *m/z* signals present in a mass spectrum, a workflow has been developed starting from an algorithm previously developed in the Cambridge University centre of Atmospheric Science group but affected by some criticism and long time consuming. The introduced modifications solved several drawbacks of the first version of the protocol, mainly due to the presence of false-positive results, and to limitation in using only the ESI source.

The compatibility with the APPI source has required the optimization of the instrumental conditions, performed by using a mixture of different classes of PAHs. The selected conditions allowed to produce sensitive measures without loss of nitro and O-PAHs by thermal decomposition.

The protocol has been used on real samples to study the capabilities of the workflow and the role of the specific ion source in describing the $PM_{2.5}$ composition. As first consideration, large data set as those produced by these experiments needs graphical tools for the data interpretation.

APPI was not suitable for the analysis of summer samples because of the absence of the photo-ionisable compounds. Usually the comparison of the applicability of ESI and

APPI refers to the known graph in Figure 46. Comparing the van Krevelen plots of all positive and negative ions detected in nanoESI and APPI, we would like to propose the comparison between the two sources in terms of elemental ratios of the ions instead than polarity. As the Figure 47 shows, nanoESI could ionize high O/C compounds and APPI low H/C substances.

.

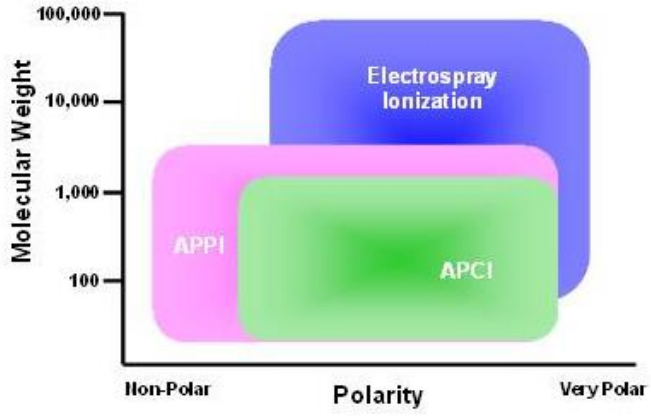

Figure 46. ESI vs. APPI based on molecular weight and polarity.

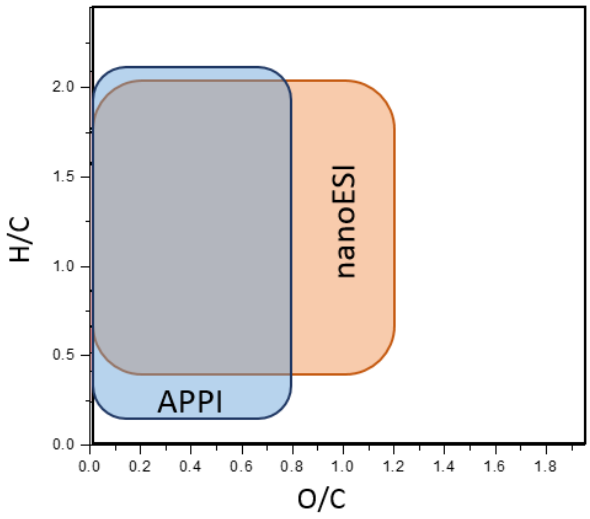

Figure 47. ESI Vs. APPI on van Krevelen plot.

Finally, we have demonstred that the overall procedure can provide a good characterization of the organic fraction of $PM_{2.5}$, providing a powerful tool for the comprehension its seasonal variability.

10 General conclusion

In this PhD thesis, the use of high-resolution mass spectrometry has been investigated for quantitative and qualitative analysis of target and non target emerging contaminants.
The ability of the HPLC-Q-TOF-MS systems to simultaneously provide conventional quantitative data in the framework of confirmatory analysis of regulated compounds and to perform a "structural-based screening" was evaluated.
The goal of this project always was the characterization of non target compounds in the same protocol of routine analysis of potential pollutants. Two different approaches were evaluated and optimized, namely a "two-steps" and autoMS analysis, consisting in the sequential collecting of MS information on molecular weights with the successive interpretation of tandem MS spectra, and in a unique protocol respectively. The most crucial parameter observed was the acquisition speeds at the MS and MS/MS levels, which affected the ability of detection of a target or non target compound.
Two methods were specifically developed for determining target and suspect cyanotoxins in freshwater intended for human consumption (chapter 6), and PDE-5 inhibitors and analogues in food supplements marked as erectile dysfunction remedies (chapter 7). The obtained results demonstrated that analytical performances requested by target analysis can be satisfactory reached in both approaches, with validation data fulfilling the criteria described in some international guidelines. In this cases in-house libraries have been created.
For the identification of compounds a system of confidence levels, from the highest 1 to the lowest 5, has been adopted, according the most recent proposal resulted from a collaborative trial on non target analysis (NORMAN project, chapter 8).
In order to confirm the structure of the target analytes, (confidence level of 1), the coupling of HRMS with a chromatographic separation, and comparison with analytical standards were necessary. In non target screening, where an analytical standard is not available the maximum identification is limited at level 2.
The application of the "two-step" and "autoMS" MS acquisitions in proper non-target-screening has been attempted in the frame of the NORMAN project described in the chapter 8.
The use of libraries is pivotal to perform this kind of analysis. The commercial availability of these libraries and the extension of the free web versions are limited.

Thus, the non-availability of specific softwares strongly limited our results; the manual peak picking and the identification of substances by interpretation of MS/MS spectra and comparison with on-line databases were extremely time-consuming. Therefore, the automatization of the data treatment seems the key for a comprehensive non-target analysis, especially in the determination of the molecular formula from the *m/z* values.
Differently, direct HRMS analysis without using chromatography, could be employed with the aim to obtain information limited to the organic composition of the sample. This approach was accomplished for the characterization of $PM_{2.5}$ organic fraction (chapter 9). In this case a specific workflow was developed improving a Mathematica based algorithm previously used by the atmospheric science group on the University of Cambridge. Results obtained overcame some issues producing false-positive results, and allowed to extend the method with the use of positive and negative APPI ionization. The workflow was successfully applied on the analysis of real sample and the study of the role of the ionization source in characterizing $PM_{2.5}$ organic fraction. All the produced data were consistent, indicating the reliability and robustness of the developed workflow.

11 Appendix

Safety

Cyanotoxins may be toxic if inhaled, absorbed through skin, or swallowed. They may cause respiratory tract, eye, and skin irritation and liver damage. Dilute solutions should be handled with laboratory gloves, eye protection, and protective clothing. If solids are weighed, this should be done with suitable ventilation and respiratory equipment.

Table S 1. List of water samples processed with related cyanobacteria identified and cyanotoxins characterized. MC stands for microcystins, ANP for anabaenopeptin.

Sample site	Sample name	Cyanobacteria occurred	Cyanotoxins identified
Apulia	Occhito, column_center_bottom 18-10-10 LW, EC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A, ANP-B, ANP-F
Apulia	Occhito, column_center_center 18-10-10 LW, EC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A
Apulia	Occhito, column_center_surface 18-10-10 LW, EC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR
Apulia	Occhito, WTP intake 18-10-10 LW, EC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR
Apulia	Occhito, WTP intake 18-10-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A, ANP-B, ANP-F, ANP-MM850, ANP-MM864
Apulia	WTP, raw 11-5-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A, ANP-B, ANP-F
Apulia	WTP, raw 1-6-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A, ANP-B, ANP-F, ANP-MM850, ANP-MM864
Apulia	WTP, raw 15-6-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, MC-LR ANP-A, ANP-B, ANP-F, ANP-

			MM850, ANP-MM864
Apulia	WTP, 15-6-10 TW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR
Apulia	WTP, raw V74-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, ANP-A, ANP-B, ANP-F, ANP-MM850, ANP-MM864, oscillammide Y
Apulia	WTP, raw V82-10 LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-F, ANP-MM850, ANP-MM864
Apulia	WTP, 15-6-11TW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR,
Apulia	WTP, raw V152, 13-12-11TW LW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	[D-Asp3] MC-RR, MC-RR, ANP-A, ANP-B, ANP-F
Apulia	WTP, V153, 13-12-11TW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	MC-RR
Apulia	WTP, V154, 13-12-11TW, TC	*Plankthotrix rubescens*, *Microcystis aeruginosa*	MC-RR
Lazio	Caprarola, raw 1A_31-3-11 LW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, ANP-A, ANP-B, ANP-F, ANP-MM850
Lazio	Caprarola, raw 27-12-11 LW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, ANP-A, ANP-B, ANP-F, ANP-MM864
Lazio	Caprarola, 2A_27-12-11 TW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR ANP-A, ANP-B,
Lazio	Caprarola, 3A_27-12-11 TW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	MC-RR, ANP-A, [D-Asp3] MC-LR
Lazio	Caprarola, raw 1A_2-1-12 LW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, ANP-A, ANP-B, ANP-MM850, ANP-MM864
Lazio	Caprarola, 3A_2-1-12 TW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	MC-RR

Lazio	Caprarola, raw 1A_27-10-12 LW, TC	*Plankthotrix rubescens*, *Aphanizomenon flos-aquae*	[D-Asp3] MC-RR, MC-RR, [D-Asp3] MC-LR, ANP-A, ANP-B, ANP-F, ANP-MM850
Sardinia	Coghinas 29-10-09 LW, TC	-	[D-Asp3] MC-RR, MC-RR MC-LR, MC-YR, ANP-A
Sardinia	Coghinas 6-10-10 LW, TC	-	MC-RR
Sardinia	Bidighinzu LW, TC	*Microcystis spp*, *Aphanizomenon spp*	ANP-A, ANP-B, ANP-F, ANP-MM850, ANP-MM865, oscillammide Y, [Asp3, MSer7] MC-LR, [MSer7] MC-LR, [MSer7] MC-YR, , MC-M(O)R, MC-FR, MC-YR, MC-LR, MC-LA, MC-LW, MC-RR, [D-Asp3] MC-RR, [D-Asp3] MC-LR, [Dha7] MC-LR, [dmAdda]MC-LR, [dmAdda] MC-YR, [D-Asp3] MC-YR

Table S 2. LC specimen of the NORMAN trial participants.

Parti.	Instrument and Model	Column	Dimensions [mm x mm, µm]	Solvent	Inj. vol [µL]
L1	AB Sciex TripleTOF 5600	Phenomenex Luna C18 (2)	2.0x150; 3.0	H_2O/ACN (FA)	10
L2	Thermo Q Exactive Orbitrap	Waters Xbridge C18	2.1x50; 3.5	H_2O/MeOH (FA)	20
L3	Agilent ToF 6230	ZIC-HILIC, Poroshell C18	2.1x150; 5.0	H_2O/ACN (NH_4Ac)	10
L4	Agilent 6550 iFunnel Q-TOF LC/MS	Waters Acquity UPLC HSS T3 and ZORBAX ECLIPSE PLUS C18 RRHD	2.1x150; 1.8 2.1x100; 1.8	H2O/MeOH (NH_4Ac)	100
L5	Waters Micromass Xevo G2 Q-TOF	Acquity UPLC BEH C18	2.1x100; 1.7	H_2O/MeOH (FA)	50
L6	AB Sciex TripleTOF 5600	Phenomenex AQ C18 (Guard), Agilent Zorbax Eclipse Plus C18	2.0x4.0; n/a	H_2O/ACN (FA)	5
L7	Waters Xevo G2-S Q-TOF	Waters Acquity HSS C18, BEH C18	2.1x150; 1.8	H_2O/ACN (NH_4fa, FA); H2O/MeOH (NH_4Ac)	5
L8	Thermo LTQ Orbitrap Discovery	Waters Atlantis HSS T3	2.1x150; 3.5	H_2O/MeOH (FA)	10
L9	Thermo LTQ Orbitrap XL	Phenomenex Kinetex C18	3.0x100; 2.6	H_2O/MeOH (FA)	10

L10	Thermo LTQ Orbitrap Discovery	Thermo Hypersil Gold	2.1x100; 3.0	H_2O/MeOH (FA)	20
L11	Agilent 6530 AM Q-TOF-LC/MS	Eclipse Plus C8	2.1x150; 3.5	H_2O/ACN (NH_4Ac, FA)	10
L12	Agilent AM Q-TOF LC/MS 6520	Phenomenex Kinetex C18	2.1x100; 1.7 and 2.1x100; 2.6	H_2O/ACN (FA, NH_4OH)	40
L13	Agilent 6550 iFunnel Q-TOF LC/MS	Zorbaax Extended-C18	2.1x50; 1.8	H_2O/ACN (FA)	2
L14a	Bruker maxis impact	Dionex Acclaim	2.1x100; 2.2	H_2O/MeOH (NH_4fa, FA, NH_4ac)	10
L15	Thermo Q Exactive Orbitrap	Hypersil Gold aQ	2.1x100; 1.9	H2O/MeOH (none, FA, NH4fa)	5
L16	Waters Xevo G2-S Q-TOF	Waters C18 BEH	2.1x100; 1.7	H_2O/MeOH (NH_4fa, pH 5)	3
L17	Agilent 6550 Q-TOF LC/MS	Agilent Poreshell HPH C18 (+Guard)	2.1x5; 2.7	H_2O/MeOH (FA pos only)	5

Table S 3. MS specimen of the NORMAN trial participants

Parti.	**Scan Range**	**Resolution ppm (*m/z*)**	**Ion source**	**Fragment Method**	**Target Software**	**Suspect, NT Procedure**
L1	100-1200	30,000 (*m/z* 400)	ESI+ only	CID, 40	Peakview, Multiquant	PeakView, Markerview. MS, isotopes, RT, MS/MS, manual peak check
L2	100-1000	140,000 (*m/z* 200)	ESI±	HCD various, 50 (DD); merged HCD 30-70 (DIA)	Xcalibur, Trace Finder, nontarget	ExactFinder 2.0, nontarget, MetFusion, internal lists. MS, RT, MS/MS, library, prediction, interpretation
L3	100-1700	12,000 (*m/z* 1000)	ESI±	In-source fragmentation at 100 V	MassHunter Qual (B.06.00)	MassHunter Qual, Profinder (B.06.00). MS, RT, internal list, MS/MS, library, STOFF-IDENT DB, pred. isotopes
L4	100-1000	56,014 (*m/z* 922)	ESI±	CID, 10, 20, 40. All Ions	MassHunter /Quant	MassHunter Quant and Qual (v B.06.00, Build 6.0.633.0). FindByFormula, MassProfiler, MS/MS, isotope, PCDL
L5	50-1000	20,000 (*m/z* 556)	ESI±	CID, 15-40 Ramp, MSE	ChromaLynx XS	ChromaLynx XS, homemade database, MassBank. MS, MSE, prediction
L6	100-1200	n/a	ESI±	CID, 40; merged 25, 40, 55. DDA	MasterView	STOFF-IDENT, DAIOS, internal list of substances. MS, (RT, MS/MS)
L7	50-1200	22,500 (*m/z* 956)	ESI±	CID, 10-55 Ramp	Unifi	Unifi. RT, MS, MS/MS, Prediction
L8	50-1000	30,000 (*m/z* 400)	ESI+ only	Not used		mzMine, ChemSpider

L9	100-1000	100,000 (*m/z* 400)	ESI±	HCD, variable	Xcalibur	Xcalibur. MS, RT, MS/MS, RT(CHI) & fragment (MetFrag) prediction, manual interpretation
L10	80-1500	30,000 (*m/z* 400)	ESI±	CID, 35	Exact Finder	Exact Finder. MS, RT, MS/MS.
L11	50-1300	n/a	ESI±	CID, 10, 20	MassHunter	Pragst library. MS, MS/MS
L12	50-2000	18,000 (*m/z* 311)	ESI±	CID, 15, 40	MassHunter Qual B04.00	Metlin database (pesticide, forensic). MS, MS/MS
L13	50-1200	20,000 (*m/z* 622)	ESI+ only	CID, 20, 40	MassHunter B6.0	ForensicsTox, Pesticides, MassBank, DAIOS. MS, RT, MS/MS, prediction
L14	50-1000	40,000 (*m/z* 1221)	ESI±	CID, 25	LCQuan, Target Analysis, Data Analysis	Target and Data Analysis. MS, isotopes, MS/MS, RT (KNN-GA-SVM), fragment (MetFrag, SmartFormLa3D) prediction
L15	70-1000	70,000 (*m/z* 200)	ESI±	HCD, 50	Trace Finder	Trace Finder, Sieve, in-house library, Thermo library (with and without RT). MS, RT, MS/MS, library & prediction
L16	50-1000	20,000 (*m/z* 556)	ESI+ only	CID, 10-45 Ramp. MSE	MassLynx,	MassLynx/ChromaLynx, in-house library. MS, RT, MS/MS
L17	50-1500	23,000 (*m/z* 119)	ESI±	CID, variable	MassHunter B06 SP1	MassHunter. MS, RT, MS/MS, prediction, Agilent ForensicTox library

Table S 4. NORMAN trial results.

Tr (min)	ion *m/z*	Ion intensity (counts)	Ion type	MS/MS available	screening approach	Proposed identification	Molecular formula	Estimated conc. [ug/l]	Identification confidence level	Retention Time Index LC-MS
10.34	412.9652	3641	M-H-	No	Target	Perfluorooctanoic Acid	$C_8HF_{15}O_2$	0.00025	Level 1	91
12.02	498.9317	22296	M-H-	Yes	Target	Perfluorooctansulfonic acid	$C_8HF_{17}O_3S$	0.001	Level 1	100.2
10.29	398.9346	22583	M-H-	Yes	Target	Perfuorohexanesulfonic acid	C6HF13O3S	0.00011	Level 1	90.8
6.69	298.9432	17734	M-H-	Yes	Target	Perfluorobutanesulfonic Acid	$C_4HF_9O_3S$	0.0036	Level 1	
11.79	171.1514	264055	M+H+	Yes	Suspect	N-[3-(dimethylamino)propyl]methacrylamide	C9H18N2O	n/a	Level 2	92.2
10.68	134.07151	30428	M+H+	Yes	Suspect	Tolyltriazole	C7H7N3	n/a	Level 2	
12.97	421.22736	2475605	M-H-	Yes	Suspect	Docusate	C20H38O7S	n/a	Level 2	102
10.09	315.25504	560049	M-H-	yes	Suspect	9,10-Dihydroxystearic acid	C18H36O4	n/a	Level 2	89.9
14.22	283.26502	1904411	M-H-	yes	Suspect	Stearic acid	C18H36O2	n/a	Level 2	104.4
15.94	429.05542	29873	M-H-	Yes	Suspect	Bicalutamid	$C_{18}H_{14}F_4N_2O_4S$	n/a	Level 2	112.8
8.11	259.1913	237602	M-H-	yes	Suspect	dihydroxymyristic acid	$C_{14}H_{28}O_4$	n/a	Level 2	
16.04	225.1963	64166	M+H+	Yes	Suspect	Dicyclohexylurea	$C_{13}H_{24}N_2O$	n/a	Level 2	107.3
8.68	246.1242	40450	M+H+	Yes	Suspect	N-Acetyl-4-aminoantipyrin	$C_{13}H_{15}N_3O_2$	n/a	Level 2	79.7
11.52	265.1477	5389191	M-H-	Yes	Suspect	dodecyl sulphate	$C_{12}H_{26}O_4S$	n/a	Level 2	98
14.89	192.1384	20240	M+H+	Yes	Suspect	N,N-diethyl-m-toluamide	$C_{12}H_{17}NO$	n/a	Level 2	103.2
28.32	391.2855	1189718	M+H+	Yes	Suspect	DEHP	$C_{24}H_{38}O_4$	n/a	Level 2	
13.17	237.104	7389	M+H+	Yes	Suspect	Carbamazepine	$C_{15}H_{12}N_2O$	n/a	Level 2	99.9
13.42	313.2387	619118	M-H-	yes	Suspect	dihydroxy-octadecenoic acid	$C_{18}H_{36}O_4$	n/a	Level 3	102.8
11.46	297.24394	744098	M-H-	Yes	Suspect	hydroxy-octadecenoic acid	$C_{18}H_{34}O_3$	n/a	Level 3	97.3
9.41	311.22349	3870768	M-H-	yes	Suspect	dihydroxy-linoleic acid	$C_{18}H_{32}O_4$	n/a	Level 3	

13.91	231.0091	8559	M-H-	No	Non-Target	Diuron	$C_9H_{10}C_{l2}N_2O$	n/a	Level 3	103.7
17.86	313.0783	18226	M-H-	No	Non-Target	Octamethyltetrasiloxane-1,7-diol	$C_8H_{26}O_5Si_4$	n/a	Level 3	
6.91	195.1231	44242	M+H+	No	Non-Target	3,6,9-trioxaundecane-1,11-diol	$C_8H_{18}O_5$	n/a	Level 3	72.8
3.2	227.9897	164385	M-H-	No	Non-Target	Picric acid	$C_6H_3N_3O_7$	n/a	Level 3	
15.82	311.29585	181369	M-H-	No	Non-Target	Arachidic acid	$C_{20}H_{40}O_2$	n/a	Level 3	112.2
12.05	301.21767	17840	M-H-	No	Non-Target	17-Methyltestosterone	$C_{20}H_{30}O_2$	n/a	Level 3	100.3
26.71	284.29549	57366	M+H+	No	Non-Target	Aldimorph	$C_{18}H_{37}NO$	n/a	Level 3	
17.87	288.2546	80180	M+H+	No	Non-Target	N,N-bis(2-hydroxyethyl)dodecanamide	$C_{16}H_{33}NO_3$	n/a	Level 3	118.8
21.46	279.161	448050	M+H+	No	Non-Target	Dibutyl Phthalate	$C_{16}H_{22}O_4$	n/a	Level 3	
21.6	279.161	448692	M+H+	No	Non-Target	Dibutyl Phthalate	$C_{16}H_{22}O_4$	n/a	Level 3	
21.45	301.1422	122471	M+Na+	No	Non-Target	Dibutyl phthalate	$C_{16}H_{22}O_4$	n/a	Level 3	
11.42	227.2019	235766	M-H-	No	Non-Target	Myristic acid	$C_{14}H_{28}O_2$	n/a	Level 3	96.9
16.08	268.15494	27711	M+H+	No	Non-Target	Diethofencarb	$C_{14}H_{21}NO_4$	n/a	Level 3	107.5
16.44	223.0973	226747	M+H+	No	Non-Target	Diethyl Phthalate (DEP)	$C_{12}H_{14}O_4$	n/a	Level 3	109.3
15.01	297.28023	138574	M-H-	No	Non-Target	nonadecanoic acid	$C_{19}H_{38}O_2$	n/a	Level 3	108.3
8.59	277.1798	47640	M-H-	No	Non-Target	Nonylphenol monocarboxylates	$C_{17}H_{26}O_3$	n/a	Level 3	
20.72	205.16003	205339	M-H-	No	Non-Target	4-Octylphenol	$C_{14}H_{22}O$	n/a	Level 3	
7.69	155.1547	256555	M+H+	Yes	Non-Target	n.i.	$C_9H_{18}N_2$	n/a	Level 4	55.8
12.45	353.2009	781446	M-H-	yes	Non-Target	n.i.	$C_{17}H_{30}N_4SO_2$	n/a	Level 4	
21.95	228.2334	483986	M+H+	Yes	Non-Target	n.i.	$C_{14}H_{29}NO$	n/a	Level 4	
8.18	323.9853	554116	M-H-	yes	Non-Target	n.i.	$C_{12}H_8ClN_3SO_4$	n/a	Level 4	
10.42	366.9003	60706	M-H-	Yes	Non-Target	n.i.	$C_{12}H_7Cl_3O_5S$	n/a	Level 4	
16.19	288.2908	26782	M+H+	Yes	Non-Target	n.i.	$C_{15}H_{35}N_4O$	n/a	Level 4	108
5.93	229.14434	127640	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	

6.09	274.1776	377247	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	
6.61	215.1258	17825	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	71.6
6.75	319.1554	153677	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	
7.38	323.9853	57252	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	
8.18	300.2026	33936	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	77.7
8.61	344.2292	35455	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	79.4
8.62	357.9464	216951	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	
9	305.187	14387	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	80.9
9.06	250.14538	50979	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	
9.22	240.1815	46874	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	81.8
9.95	520.335	16610	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	84.6
10.74	163.1328	53265	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	87.7
10.95	294.1565	21497	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	88.6
12.19	312.2388	52011	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	93.9
13.14	488.3448	10736	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	99.8
14	289.2024	149802	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	101.5
15.09	251.1688	18762	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	103.6
15.73	214.0908	86310	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	105.8
17.21	209.1903	179320	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	113.1
17.33	244.2291	70125	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	113.9
17.69	159.1386	54629	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	117.2
17.84	332.2811	137633	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	118.6
18.16	219.1749	49318	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	121.5
19.19	335.2208	70281	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	129
19.75	294.2076	84046	M+NH4+	No	Non-Target	n.i.	n/a	n/a	Level 5	132.6

20.29	235.1701	155995	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	128.3
20.53	294.2076	238839	M+NH4+	No	Non-Target	n.i.	n/a	n/a	Level 5	141.3
20.88	203.1076	57952	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	
22.41	556.4436	12576	M+H+	No	Non-Target	n.i.	n/a	n/a	Level 5	
12.82	351.21938	143645	M-H-	No	Non-Target	n.i.	n/a	n/a	Level 5	101.7
6.61	157.1231	219748	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	
7.55	114.0918	68956	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	
7.87	329.2333	361497	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	
8.03	329.2333	385396	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	
8.23	329.2333	159300	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	
8.28	329.2333	168736	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	
9.01	388.2559	32238	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	81
9.3	340.2612	72839	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	82.1
9.35	432.2825	26596	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	82.3
10.23	277.1448	1259183	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	90.5
11.49	367.1919	1157239	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	97.7
11.76	279.1055	965349	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	99.7
12.49	356.266	46281	M+NH4+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	95.2
12.75	400.2923	31190	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	97.3
12.77	293.1796	6234448	M-H-	Yes	Non-Target	n.i.	n/a	n/a	Level 5	101.6
12.91	485.2818	221385	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	101.9
13.32	485.2818	239400	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	102.6
13.38	337.2054	541625	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	102.7
14.2	557.3341	82826	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	104.3
14.53	365.2374	836358	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	105.9

14.59	341.2704	445131	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	106.2
15.03	453.29	538286	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	108.4
15.06	349.2425	517037	M-H-	yes	Non-Target	n.i.	n/a	n/a	Level 5	108.5
16.03	465.30471	136609	M-H-	Yes	Non-Target	n.i.	n/a	n/a	Level 5	113.3
16.78	209.1546	327117	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	111
19.18	330.2647	94506	M+H+	Yes	Non-Target	n.i.	n/a	n/a	Level 5	128.9

Table S 5. PAHs standard solution concentration.

Compound	Molecular formula	Stock solution (µg/mL)	Concentration after extraction (ng/mL)	theoretical concentration in PM (µg/m3)
Acenaphthene	$C_{12}H_{10}$	333	2222	485
Acenaphthylene	$C_{12}H_{8}$	667	4444	970
Anthracene	$C_{14}H_{10}$	33	222	48
Benz[a]anthracene	$C_{18}H_{12}$	33	222	48
Benzo[b]fluoranthene	$C_{20}H_{12}$	33	222	48
Benzo[k]fluoranthene	$C_{20}H_{12}$	33	222	48
Benzo[ghi]perylene	$C_{22}H_{12}$	67	444	97
Benzo[a]pyrene	$C_{20}H_{12}$	33	222	48
Chrysene	$C_{18}H_{12}$	33	222	48
Dibenz[a,h]anthracene	$C_{22}H_{14}$	67	444	97
Fluoranthene	$C_{16}H_{10}$	33	222	48
Fluorene	$C_{13}H_{1}0$	67	444	97
Indeno[1,2,3-cd]pyrene	$C_{22}H_{12}$	33	222	48
1-Methylnaphthalene	$C_{11}H_{10}$	333	2222	485
2-Methylnaphthalene	$C_{11}H_{10}$	333	2222	485
Naphthalene	$C_{10}H_{8}$	333	2222	485
Phenanthrene	$C_{14}H_{10}$	33	222	48
Pyrene	$C_{16}H_{10}$	33	222	48
9-nitroanthracene	$C_{14}H_{9}NO_{2}$	3	22	5
4-nitrocatechol	$C_{6}H_{5}NO_{4}$	27	178	39
4-nitrophenol	$C_{6}H_{5}NO_{3}$	13	89	19
9,10-antraquinone	$C_{14}H_{8}O_{2}$	67	444	97
9-phenanthrenecarboxaldehyde	$C_{15}H_{10}O$	13	89	19
9-fluorenone	$C_{13}H_{8}O$	13	89	19
1-naphthaldeyde	$C_{11}H_{8}O$	27	178	39
9-hydroxyphenanthrene	$C_{14}H_{10}O$	0.65	4	1
9-hydroxyfluorene	$C_{13}H_{10}O$	0.65	4	1

Table S 6. Results for the dopant optimization.

				Toluene 10%		Toluene 5%		Acetone 10%		Acetone 5%	
Compound	**Molecular Formula**	**Ion type**	***m/z***	**Intensity**	**Ion type ratio**	**Intensity**	**Ion type ratio**	**Intensity**	**Ion type ratio**	**Intensity**	**Ion type ratio**
Acenaphthene	$C_{12}H_{10}$	M+H+	155.0855	7.4E+06	4%	6.0E+06	6%	2.3E+05	3%	2.8E+05	3%
		M+	154.0772	1.7E+08	96%	1.0E+08	94%	7.6E+06	97%	9.1E+06	97%
Acenaphthylene	$C_{12}H_8$	M+H+	153.0699	6.6E+07	33%	4.6E+07	37%	3.2E+06	42%	4.0E+06	44%
		M+	152.0615	1.4E+08	67%	8.0E+07	63%	4.5E+06	58%	5.1E+06	56%
Anthracene Phenanthrene	$C_{14}H_{10}$	M+H+	179.0855	9.6E+06	19%	5.8E+06	23%	2.4E+06	47%	2.7E+06	48%
		M+	178.0772	4.0E+07	81%	1.9E+07	77%	2.7E+06	53%	3.0E+06	52%
Benz[a]anthracene +Chrysene	$C_{18}H_{12}$	M+H+	229.1012	4.7E+06	21%	3.9E+06	27%	1.2E+05	11%	3.0E+05	15%
		M+	228.0928	1.7E+07	79%	1.0E+07	73%	9.7E+05	89%	1.7E+06	85%
Benzo[b]fluoranthene Benzo[k]fluoranthene Benzo[a]pyrene	$C_{20}H_{12}$	M+H+	253.1012	2.4E+06	29%	3.1E+06	44%	3.9E+05	59%	1.2E+06	62%
		M+	252.0928	5.9E+06	71%	4.0E+06	56%	2.7E+05	41%	7.2E+05	38%
Benzo[ghi]perylene Indeno[1,2,3-cd]pyrene	$C_{22}H_{12}$	M+H+	277.1012	1.8E+06	46%	6.7E+05	54%	1.1E+05	50%	1.9E+05	54%
		M+	276.0928	2.0E+06	54%	5.8E+05	46%	1.0E+05	50%	1.6E+05	46%
Dibenz[a,h]anthracene	$C_{22}H_{14}$	M+H+	279.1168	9.9E+05	46%	5.4E+05	56%	9.7E+04	59%	1.6E+05	57%
		M+	278.1085	1.2E+06	54%	4.2E+05	44%	6.6E+04	41%	1.2E+05	43%
Fluoranthene+ Pyrene	$C_{16}H_{10}$	M+H+	203.0855	6.5E+06	13%	3.2E+06	17%	1.3E+05	5%	2.5E+05	8%
		M+	202.0772	4.2E+07	87%	1.6E+07	83%	2.4E+06	95%	3.1E+06	92%
Fluorene	$C_{13}H_{10}$	M+H+	167.0855	5.1E+05	2%	3.3E+05	2%	1.8E+05	16%	1.7E+05	13%
		M+	166.0772	2.7E+07	98%	1.6E+07	98%	9.7E+05	84%	1.1E+06	87%
1-Methylnaphthalene 2-	$C_{11}H_{10}$	M+H+	143.0855	6.7E+05	0%	5.0E+05	0%	4.1E+05	7%	3.4E+05	5%

Methylnaphthalene		M+	142.0772	1.6E+08	100%	1.0E+08	100%	5.8E+06	93%	6.6E+06	95%
Naphthalene	$C_{10}H_8$	M+H+	129.0699	7748663	13%	6.1E+06	16%	3.1E+06	35%	3.2E+06	36%
		M+	128.0621	5.0E+07	87%	3.1E+07	84%	5.6E+06	65%	5.8E+06	64%
9-nitroanthracene	$C_{14}H_9NO_2$	M+H+	224.0706	4.1E+06	24%	2.2E+06	30%	9.6E+04	18%	2.0E+05	25%
		M+	223.0622	1.3E+07	76%	5.2E+06	70%	4.4E+05	82%	6.0E+05	75%
4-nitrocatechol	$C_6H_5NO_4$	M+H+	156.0291	9.9E+05	100%	1.3E+06	100%	/		/	
		M+	155.0167		0%		0%	/		/	
4-nitrophenol	$C_6H_5NO_3$	M+H+	140.0342	7.2E+06	100%	4.2E+06	100%	2.2E+05	100%	2.8E+05	100%
		M+	139.0258		0%		0%		0%		0%
9,10-antraquinone	$C_{14}H_8O_2$	M+H+	209.0597	6.0E+07	100%	2.7E+07	100%	5.7E+06	100%	8.3E+06	100%
		M+	208.0513		0%		0%		0%		0%
9-phenanthrene carboxaldehyde	$C_{15}H_{10}O$	M+H+	207.0804	1.9E+06	100%	4.0E+06	100%	6.4E+06	100%	6.8E+06	100%
		M+	206.0721		0%		0%		0%		0%
9-fluorenone	$C_{13}H_8O$	M+H+	181.0648	3.4E+07	98%	2.3E+07	98%	3.3E+07	100%	3.0E+07	100%
		M+	180.0564	7.9E+05	2%	3.5E+05	2%		0%		0%
1-naphthaldeyde	$C_{11}H_8O$	M+H+	157.0648	1.8E+07	98%	1.4E+07	98%	7.6E+06	98%	7.1E+06	98%
		M+	156.0564	4.2E+05	2%	2.9E+05	2%	1.8E+05	2%	1.6E+05	2%
9-hydroxy phenanthrene	$C_{14}H_{10}O$	M+H+	195.0804	2.6E+06	61%	1.5E+06	67%	3.2E+06	80%	2.7E+06	79%
		M+	194.0721	1.7E+06	39%	7.5E+05	33%	8.1E+05	20%	7.4E+05	21%
9-hydroxyfluorene	$C_{13}H_{10}O$	M+H+	183.0804	1.8E+07	98%	1.6E+07	99%	7.1E+06	100%	7.6E+06	100%
		M+	182.0721	3.7E+05	2%	1.8E+05	1%		0%		0%

Table S 7. Results for the extraction optimization

Compound Name	Neutral mulecular formula	Ion type	*m/z*	St. Mix Meth.	Spiked sample Meth.	Spiked sample Meth. Recons.	St. Mix Meth: CH2Cl2	Spiked sample Meth: CH2Cl2	Spiked sample Meth: CH2Cl2 Recons.	Ratio St. Mix Meth/ CH2Cl 2: Meth.	Recovery Methanol extraction	Recovery CH2Cl2: Methanol extraction	Recovery Methanol Recons. extraction	Recovery Methanol: CH2Cl2 Recons. extraction	Ratio Sample extraction Meth/CH2 Cl2: Meth
Acenaphthene	$C_{12}H_{10}$	M+H+	155.0855	82929	48295	0	70118	71484	23315	118%	58%	102%	0%	33%	68%
		M+	154.0772	1040393	584012	10090	844771	1138000	18302	123%	56%	135%	1%	2%	51%
Acenaphthylene	$C_{12}H_8$	M+H+	153.0699	613343	348090	6950	428402	485977	12275	143%	57%	113%	1%	3%	72%
		M+	152.0615	1685338	960778	4440	1075502	1028542	8640	157%	57%	96%	0%	1%	93%
Anthracene Phenanthrene	$C_{14}H_{10}$	M+H+	179.0855	61363	57703	0	112315	60663	20038	55%	94%	54%	0%	18%	95%
		M+	178.0772	172274	143521	19470	274893	202536	29157	63%	83%	74%	11%	11%	71%
Benz[a]anthracene +Chrysene	$C_{18}H_{12}$	M+H+	229.1012	5374	25314	6377	24742	0	0	22%	471%	0%	119%	0%	
		M+	228.0928	13896	23868	10676	42349	11516	20025	33%	172%	27%	77%	47%	207%
Benzo[b]fluoranthene Benzo[k]fluoranthene Benzo[a]pyrene	$C_{20}H_{12}$	M+H+	253.1012	13552	18728	14040	46403	15456	20700	29%	138%	33%	104%	45%	121%
		M+	252.0928	14478	23381	22999	41950	21608	39111	35%	161%	52%	159%	93%	108%
Benzo[ghi]perylene Indeno[1,2,3-cd]pyrene	$C_{22}H_{12}$	M+H+	277.1012	13023	22754	22446	66700	20029	0	20%	175%	30%	172%	0%	114%
		M+	276.0928	0	0	0	14298	0	0			0%		0%	
Dibenz[a,h]anthracene	$C_{22}H_{14}$	M+H+	279.1168	5866	6992	7257	9558	0	0	61%	119%	0%	124%	0%	
		M+	278.1085	0	0	0	0	0	0						
Fluoranthene+ Pyrene	$C_{16}H_{10}$	M+H+	203.0855	28416	34615	15067	56651	16557	30521	50%	122%	29%	53%	54%	209%
		M+	202.0772	83891	108750	46261	166011	59140	90392	51%	130%	36%	55%	54%	184%
Fluorene	$C_{13}H_{10}$	M+H+	167.0855	0	9373	0	16834	13997	0			83%		0%	67%
		M+	166.0772	185735	125659	0	214143	244777	6864	87%	68%	114%	0%	3%	51%
1-Methylnaphthalene 2-Methylnaphthalene	$C_{11}H_{10}$	M+H+	143.0855	35780	24920	21593	49763	43789	40427	72%	70%	88%	60%	81%	57%
		M+	142.0772	2034504	1107084	0	1612386	1274586		126%	54%	79%	0%	0%	87%
Naphthalene	$C_{10}H_8$	M+H+	129.0699	393619	204944	75217	398525	299923	101430	99%	52%	75%	19%	25%	68%
		M+	128.062	1302392	515630	84653	469205	645781	142297	278%	40%	138%	6%	30%	80%

9-nitroanthracene	$C_{14}H_9NO_2$	M+H+	224.0706	0	0	0	0	0	0								
		M+	223.0622	0	0	0	0	0	0								
4-nitrocatechol	$C_6H_5NO_4$	M+H+	156.0291	0	0	0	0	0	0								
		M+	155.0167	0	0	0	0	0	0								
4-nitrophenol	$C_6H_5NO_3$	M+H+	140.0342	0	0	0	0	0	0								
		M+	139.0258	0	0	0	0	0	0								
9,10-antraquinone	$C_{14}H_8O_2$	M+H+	209.0597	352535	500340	268888	799886	226127	563176	44%	142%	28%	76%	70%	221%		
		M+	208.0513	0	0	0	0	0	0								
9-phenanthrene carboxaldehyde	$C_{15}H_{10}O$	M+H+	207.0804	63732	76569	41413	104284	14032	45029	61%	120%	13%	65%	43%	546%		
		M+	206.0721	0	0	0	0	0	0								
9-fluorenone	$C_{13}H_8O$	M+H+	181.0648	823283	583312	241804	1135924	465807	211591	72%	71%	41%	29%	19%	125%		
		M+	180.0564	28708	18004	15708	35445	21518		81%	63%	61%	55%	0%	84%		
1-naphthaldeyde	$C_{11}H_8O$	M+H+	157.0648	595681	297224	41835	546269	422489	26017	109%	50%	77%	7%	5%	70%		
		M+	156.0564	0	2160	0	0	0	0								
9-hydroxy phenanthrene	$C_{14}H_{10}O$	M+H+	195.0804	25558	16245	9547	35596	16862	12055	72%	64%	47%	37%	34%	96%		
		M+	194.0721	0	1482	0	0	0	0								
9-hydroxyfluorene	$C_{13}H_{10}O$	M+H+	183.0804	342332	180662	0	320043	189506	113274	107%	53%	59%	0%	35%	95%		
		M+	182.0721	0	0	0	0	0	0								
				Number of detected compounds						AVERAGE							
				26	29	21	28	25	21	83%	105%	60%	47%	25%	122%		

12 References

[1] D.H. Chace, Mass spectrometry in newborn and metabolic screening: Historical perspective and future directions, J. Mass Spectrom. 44 (**2009**) 163-170.

[2] J.E. Drewes, L.S. Shore, Concerns about pharmaceuticals in water reuse, groundwater recharge, and animal waste. **2001**.

[3] H. Fromme, S.A. Tittlemier, W. Völkel, M. Wilhelm, D. Twardella, Perfluorinated compounds - Exposure assessment for the general population in western countries, Int. J. Hyg. Environ. Health 212 (2009) 239-270.

[4] C.A. De Wit, An overview of brominated flame retardants in the environment, Chemosphere 46 (**2002**) 583-624.

[5] G. Oberdörster, E. Oberdörster, J. Oberdörster, Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles, Environ. Health Perspect. 113 (**2005**) 823-839.

[6] F.T. Lange, M. Scheurer, H.-. Brauch, Artificial sweeteners-A recently recognized class of emerging environmental contaminants: A review, Anal. Bioanal. Chem. 403 (**2012**) 2503-2518.

[7] M.l. Farré, S. Pérez, L. Kantiani, D. Barceló, Fate and toxicity of emerging pollutants, their metabolites and transformation products in the aquatic environment, TrAC Trends Anal. Chem. 27 (**2008**) 991-1007.

[8] S.D. Richardson, T.A. Ternes, Water analysis: Emerging contaminants and current issues, Anal. Chem. 86 (**2014**) 2813-2848.

[9] S. Merel, D. Walker, R. Chicana, S. Snyder, E. Baurès, O. Thomas, State of knowledge and concerns on cyanobacterial blooms and cyanotoxins, Environ. Int. 59 (**2013**) 303-327.

[10] R.M. Dawson, The toxicology of microcystins, Toxicon 36 (**1998**) 953-962.

[11] A. Fristachi, J.L. Sinclair, S. Hall, J.A.H. Berkman, G. Boyer, J. Burkholder, J. Burns, W. Carmichael, A. Dufour, W. Frazier, S.L. Morton, E. O'Brien, S. Walker, Occurrence of cyanobacterial harmful algal blooms workgroup report. **2008**.

[12] S. Bogialli, F. Nigro Di Gregorio, L. Lucentini, E. Ferretti, M. Ottaviani, N. Ungaro, P.P. Abis, M. Cannarozzi De Grazia, Management of a toxic cyanobacterium bloom (Planktothrix rubescens) affecting an Italian drinking water basin: A case study, Environ. Sci. Technol. 47 (**2013**) 574-583.

[13] R. El-Shehawy, E. Gorokhova, F. Fernández-Piñas, F.F. del Campo, Global warming and hepatotoxin production by cyanobacteria: What can we learn from experiments? Water Res. 46 (2012) 1420-1429.

[14] H.W. Paerl, V.J. Paul, Climate change: Links to global expansion of harmful cyanobacteria, Water Res. 46 (**2012**) 1349-1363.

[15] J. Heisler, P.M. Glibert, J.M. Burkholder, D.M. Anderson, W. Cochlan, W.C. Dennison, Q. Dortch, C.J. Gobler, C.A. Heil, E. Humphries, A. Lewitus, R. Magnien, H.G. Marshall, K. Sellner, D.A. Stockwell, D.K. Stoecker, M. Suddleson, Eutrophication and harmful algal blooms: A scientific consensus, Harmful Algae 8 (**2008**) 3-13.

[16] J.A. Downing, S.B. Watson, E. McCauley, Predicting cyanobacteria dominance in lakes, Can. J. Fish. Aquatic Sci. 58 (**2001**) 1905-1908.

[17] W.W. Carmichael, Cyanobacteria secondary metabolites - The cyanotoxins, J. Appl. Bacteriol. 72 (**1992**) 445-459.

[18] M.E. Van Apeldoorn, H.P. Van Egmond, G.J.A. Speijers, G.J.I. Bakker, Toxins of cyanobacteria, Mol. Nutr. Food Res. 51 (**2007**) 7-60.

[19] V. Messineo, S. Bogialli, S. Melchiorre, N. Sechi, A. Lugliè, P. Casiddu, M.A. Mariani, B.M. Padedda, A.D. Corcia, R. Mazza, E. Carloni, M. Bruno, Cyanobacterial toxins in Italian freshwaters, Limnologica 39 (**2009**) 95-106.

[20] S. Gkelis, V. Harjunpää, T. Lanaras, K. Sivonen, Diversity of hepatotoxic microcystins and bioactive anabaenopeptins in cyanobacterial blooms from Greek freshwaters, Environ. Toxicol. 20 (**2005**) 249-256.

[21] U. Neumann, A. Forchert, T. Flury, J. Weckesser, Microginin FR1, a linear peptide from a water bloom of Microcystis species, FEMS Microbiol. Lett. 153 (**1997**) 475-478.

[22] K. Ishida, T. Kato, M. Murakami, M. Watanabe, M.F. Watanabe, Microginins, zinc metalloproteases inhibitors cyanobacterium Microcystis aeruginosa, Tetrahedron 56 (**2000**) 8643-8656.

[23] E. Funari, E. Testai, Human health risk assessment related to cyanotoxins exposure, Crit. Rev. Toxicol. 38 (**2008**) 97-125.

[24] W.W. Carmichael, S.M.F.O. Azevedo, J.S. An, R.J.R. Molica, E.M. Jochimsen, S. Lau, K.L. Rinehart, G.R. Shaw, G.K. Eaglesham, Human fatalities form cyanobacteria: Chemical and biological evidence for cyanotoxins, Environ. Health Perspect. 109 (**2001**) 663-668.

[25] W. Word Health Organization, Guidelines for drinking water quality, third edition, (**2013**) .

[26] B.J. Venhuis, D. de Kaste, Towards a decade of detecting new analogues of sildenafil, tadalafil and vardenafil in food supplements: A history, analytical aspects and health risks, J. Pharm. Biomed. Anal. 69 (**2012**) 196-208.

[27] O. Corazza, G. Martinotti, R. Santacroce, E. Chillemi, M. Di Giannantonio, F. Schifano, S. Cellek, Sexual enhancement products for sale online: Raising awareness of the psychoactive effects of yohimbine, maca, horny goat weed, and ginkgo biloba, BioMed Research International 2014 (**2014**) .

[28] R. ShamLoul, Natural aphrodisiacs, Journal of Sexual Medicine 7 (**2010**) 39-49.

[29] E. Wespes, E. Amar, D. Hatzichristou, K. Hatzimouratidis, F. Montorsi, J. Pryor, Y. Vardi, EAU Guidelines on Erectile Dysfunction: An Update, Eur. Urol. 49 (**2006**) 806-815.

[30] European Parliament and Council, amending, as regards traditional herbal medicinal products, Directive 2001/83/EC on the Community code relating to medicinal products for human use, OJ L 136 (**2004**) 85-90.

[31] D.N. Patel, L. Li, C.-. Kee, X. Ge, M.-. Low, H.-. Koh, Screening of synthetic PDE-5 inhibitors and their analogues as adulterants: Analytical techniques and challenges, J. Pharm. Biomed. Anal. 87 (**2014**) 176-190.

[32] N. Schramek, U. Wollein, W. Eisenreich, Identification of new synthetic PDE-5 inhibitors analogues found as minor components in a dietary supplement, J. Pharm. Biomed. Anal. 96 (**2014**) 45-53.

[33] Y.-. Lam, W.-. Poon, C.-. Lai, A.Y.-. Chan, T.W.-. Mak, Identification of a novel vardenafil analogue in herbal product, J. Pharm. Biomed. Anal. 46 (**2008**) 804-807.

[34] J. Vaysse, V. Gilard, S. Balayssac, C. Zedde, R. Martino, M. Malet-Martino, Identification of a novel sildenafil analogue in an adulterated herbal supplement, J. Pharm. Biomed. Anal. 59 (**2012**) 58-66.

[35] V.M. Toomey, J.J. Litzau, C.L. Flurer, Isolation and structural characterization of two tadalafil analogs found in dietary supplements, J. Pharm. Biomed. Anal. 59 (**2012**) 50-57.

[36] M.-. Shin, M.-. Hong, W.-. Kim, Y.-. Lee, Y.-. Jeoung, Identification of a new analogue of sildenafil added illegally to a functional food marketed for penile erectile dysfunction, Food Addit. Contam. 20 (**2003**) 793-796.

[37] S. Singh, B. Prasad, A.A. Savaliya, R.P. Shah, V.M. Gohil, A. Kaur, Strategies for characterizing sildenafil, vardenafil, tadalafil and their analogues in herbal dietary supplements, and detecting counterfeit products containing these drugs, TrAC - Trends in Analytical Chemistry 28 (**2009**) 13-28.

[38] The European parliament and the council of the European Union., DIRECTIVE 2002/46/EC , Off. J. Eur. Commun L 183 (**2002**) 51-57.

[39] Organization for Economic Co-operation and Development (OECD); OECD ENVIRONMENTAL OUTLOOK TO 2050: The Consequences of Inaction, (**2012**) .

[40] E. Von Schneidemesser, P.S. Monks, J.D. Allan, L. Bruhwiler, P. Forster, D. Fowler, A. Lauer, W.T. Morgan, P. Paasonen, M. Righi, K. Sindelarova, M.A. Sutton, Chemistry and the Linkages between Air Quality and Climate Change, Chem. Rev. 115 (**2015**) 3856-3897.

[41] M. Hallquist, J.C. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, J. Dommen, N.M. Donahue, C. George, A.H. Goldstein, J.F. Hamilton, H. Herrmann, T. Hoffmann, Y. Iinuma, M. Jang, M.E. Jenkin, J.L. Jimenez, A. Kiendler-Scharr, W. Maenhaut, G. McFiggans, T.F. Mentel, A. Monod, A.S.H. Prévôt, J.H. Seinfeld, J.D. Surratt, R. Szmigielski, J. Wildt, The formation, properties and impact of secondary organic aerosol: Current and emerging issues, Atmos. Chem. Phys. 9 (**2009**) 5155-5236.

[42] B. Nozière, M. Kalberer, M. Claeys, J. Allan, B. D'Anna, S. Decesari, E. Finessi, M. Glasius, I. Grgic, J.F. Hamilton, T. Hoffmann, Y. Iinuma, M. Jaoui, A. Kahnt, C.J. Kampf, I. Kourtchev, W. Maenhaut, N. Marsden, S. Saarikoski, J. Schnelle-Kreis, J.D. Surratt, S. Szidat, R. Szmigielski, A. Wisthaler, The Molecular Identification of Organic Compounds in the Atmosphere: State of the Art and Challenges, Chem. Rev. 115 (**2015**) 3919-3983.

[43] J.H. Kroll, J.H. Seinfeld, Chemistry of secondary organic aerosol: Formation and evolution of low-volatility organics in the atmosphere, Atmos. Environ. 42 (**2008**) 3593-3624.

[44] R. Zhang, G. Wang, S. Guo, M.L. Zamora, Q. Ying, Y. Lin, W. Wang, M. Hu, Y. Wang, Formation of Urban Fine Particulate Matter, Chem. Rev. 115 (**2015**) 3803-3855.

[45] A.G. Rincón, A.I. Calvo, M. Dietzel, M. Kalberer, Seasonal differences of urban organic aerosol composition an ultra-high resolution mass spectrometry study, Environ. Chem. 9 (**2012**) 298-319.

[46] S.A. Nizkorodov, J. Laskin, A. Laskin, Molecular chemistry of organic aerosols through the application of high resolution mass spectrometry, Phys. Chem. Chem. Phys. 13 (**2011**) 3612-3629.

[47] Pope III C.A., Dockery D.W., Health effects of fine particulate air pollution: Lines that connect, J. Air Waste Manage. Assoc. 56 (**2006**) 709-742.

[48] M.L. Bell, D.L. Davis, Reassessment of the lethal London fog of 1952: Novel indicators of acute and chronic consequences of acute exposure to air pollution, Environ. Health Perspect. 109 (**2001**) 389-394.

[49] Chow J.C., Watson J.G., Zones of representation for PM10 measurements along the US/Mexico border, Sci. Total Environ. 276 (**2001**) 49-68.

[50] He K., Yang F., Ma Y., Zhang Q., Yao X., Chan C.K., Cadle S., Chan T., Mulawa P., The characteristics of PM2.5 in Beijing, China, Atmos. Environ. 35 (**2001**) 4959-4970.

[51] J. Löndahl, A. Massling, J. Pagels, E. Swietlicki, E. Vaclavik, S. Loft, Size-resolved respiratory-tract deposition of fine and ultrafine hydrophobic and hygroscopic aerosol particles during rest and exercise, Inhal. Toxicol. 19 (**2007**) 109-116.

[52] K.-. Kim, E. Kabir, S. Kabir, A review on the human health impact of airborne particulate matter, Environ. Int. 74 (**2015**) 136-143.

[53] W. Fang, Y. Yang, Z. Xu, PM10 and PM2.5 and health risk assessment for heavy metals in a typical factory for cathode ray tube television recycling, Environ. Sci. Technol. 47 (**2013**) 12469-12476.

[54] F. Peter Guengerich, Metabolic activation of carcinogens, Pharmacol. Ther. 54 (**1992**) 17-61.

[55] W. Xue, D. Warshawsky, Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review, Toxicol. Appl. Pharmacol. 206 (**2005**) 73-93.

[56] Twomey S.A., Piepgrass M., Wolfe T.L., An assessment of the impact of pollution on global cloud albedo, Tellus 36 B (**1984**) 356-366.

[57] V.P.S. Almeida, K. Cogo, S.M. Tsai, D.H. Moon, Colorimetric test for the monitoring of microcystins in cyanobacterial culture and environmental samples from southeast - Brazil, Braz. J. Microbiol. 37 (**2006**) 192-198.

[58] J. Al-Tebrineh, M.M. Gehringer, R. Akcaalan, B.A. Neilan, A new quantitative PCR assay for the detection of hepatotoxigenic cyanobacteria, Toxicon 57 (**2011**) 546-554.

[59] Johnson C.M., Pleshko N., Achary M., Suri R.P., 38th Annual Northeast Bioengineering Conference, NEBEC (**2012)**, 57.

[60] Meneely J.P., Ricci F., van Egmond H.P., Elliott C.T., Current methods of analysis for the determination of trichothecene mycotoxins in food, TrAC Trends Anal. Chem. 30 (**2011**) 192-203.

[61] Lachenmeier D.W., Eberhard H., Fang F., Birk S., Peter D., Constanze S., Manfred S., NMR-spectroscopy for nontargeted screening and simultaneous quantification of health-relevant compounds in foods: The example of melamine, J. Agric. Food Chem. 57 (**2009**) 7194-7199.

[62] Anastassiades M., Lehotay S.J., Štajnbaher D., Schenck F.J., Fast and easy multiresidue method employing acetonitrile extraction/partitioning and "dispersive solid-phase extraction" for the determination of pesticide residues in produce, J. AOAC Int. 86 (**2003**) 412-431.

[63] Ibáñez M., Sancho J.V., Pozo Ó.J., Niessen W., Hernández F., Use of quadrupole time-of-flight mass spectrometry in the elucidation of unknown compounds present in environmental water, Rapid Commun. Mass Spectrom. 19 (**2005**) 169-178.

[64] Grimalt S., Pozo Ó.J., Sancho J.V., Hernández F., Use of liquid chromatography coupled to quadrupole time-of-flight mass spectrometry to investigate pesticide residues in fruits, Anal. Chem. 79 (**2007**) 2833-2843.

[65] E. Beltrán, M. Ibáñez, J.V. Sancho, F. Hernández, Determination of six microcystins and nodularin in surface and drinking waters by on-line solid phase extraction-ultra high pressure liquid chromatography tandem mass spectrometry, J. Chromatogr. A 1266 (**2012**) 61-68.

[66] Zachariasova M., Lacina O., Malachova A., Kostelanska M., Poustka J., Godula M., Hajslova J., Novel approaches in analysis of Fusarium mycotoxins in cereals employing ultra performance liquid chromatography coupled with high resolution mass spectrometry, Anal. Chim. Acta 662 (**2010**) 51-61.

[67] M. Farré, L. Kantiani, M. Petrovic, S. Pérez, D. Barceló, 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**) 86-99.

[68] 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**) 510-519.

[69] M. Ibáñez, J.V. Sancho, F. Hernández, D. McMillan, R. Rao, Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-light mass spectrometry, TrAC Trends Anal. Chem. 27 (**2008**) 481-489.

[70] J. Nurmi, J. Pellinen, A.-. 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**) 303-312.

[71] M. Zedda, C. Zwiener, Is nontarget screening of emerging contaminants by LC-HRMS successful? A plea for compound libraries and computer tools, Anal. Bioanal. Chem. 403 (**2012**) 2493-2502.

[72] M.-. Neffling, L. Spoof, M. Quilliam, J. Meriluoto, LC-ESI-Q-TOF-MS for faster and accurate determination of microcystins and nodularins in serum, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 878 (**2010**) 2433-2441.

[73] P. Ferranti, S. Fabbrocino, E. Chiaravalle, M. Bruno, A. Basile, L. Serpe, P. Gallo, Profiling microcystin contamination in a water reservoir by MALDI-TOF and liquid chromatography coupled to Q/TOF tandem mass spectrometry, Food Res. Int. 54 (**2013**) 1321-1330.

[74] E.L. Schymanski, H.P. Singer, J. Slobodnik, I.M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, N.S. Thomaidis, A. Bletsou, C. Zwiener, M. Ibáñez, T. Portolés, R. de Boer, M.J. Reid, M. Onghena, U. Kunkel, W. Schulz, A. Guillon, N. Noyon, G. Leroy, P. Bados, S. Bogialli, D. Stipanicev, P. Rostkowski, J. Hollender, Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis, Anal. Bioanal. Chem.(**2015**) .

[75] 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. 48 (**2014**) 2097-2098.

[76] EUROPEAN COMMUNITY, GUIDELINES FOR THE IMPLEMENTATION OF DECISION 2002/657/EC 1 REGARDING SOME CONTAMINANTS, (**2007**) .

[77] E.L. Schymanski, H.P. Singer, P. Longrée, M. Loos, M. Ruff, M.A. Stravs, C. Ripollés Vidal, J. Hollender, Strategies to characterize polar organic contamination in wastewater: Exploring the capability of high resolution mass spectrometry, Environ. Sci. Technol. 48 (**2014**) 1811-1818.

[78] T. Reemtsma, Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry. Status and needs, J. Chromatogr. A 1216 (**2009**) 3687-3701.

[79] S. Kim, R.W. Kramer, P.G. Hatcher, Graphical Method for Analysis of Ultrahigh-Resolution Broadband Mass Spectra of Natural Organic Matter, the Van Krevelen Diagram, Anal. Chem. 75 (**2003**) 5336-5344.

[80] T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinform. 8 (**2007**) .

[81] Wu Z., Rodgers R.P., Marshall A.G., Two- and Three-Dimensional van Krevelen Diagrams: A Graphical Analysis Complementary to the Kendrick Mass Plot for Sorting Elemental Compositions of

Complex Organic Mixtures Based on Ultrahigh-Resolution Broadband Fourier Transform Ion Cyclotron Resonance Mass Measurements, Anal. Chem. 76 (**2004**) 2511-2516.

[82] J.H. Kroll, N.M. Donahue, J.L. Jimenez, S.H. Kessler, M.R. Canagaratna, K.R. Wilson, K.E. Altieri, L.R. Mazzoleni, A.S. Wozniak, H. Bluhm, E.R. Mysak, J.D. Smith, C.E. Kolb, D.R. Worsnop, Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol, Nat. Chem. 3 (**2011**) 133-139.

[83] Hughey C.A., Hendrickson C.L., Rodgers R.P., Marshall A.G., Qian K., Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra, Anal. Chem. 73 (**2001**) 4676-4681.

[84] I. Kourtchev, I.P. O'Connor, C. Giorio, S.J. Fuller, K. Kristensen, W. Maenhaut, J.C. Wenger, J.R. Sodeau, M. Glasius, M. Kalberer, Effects of anthropogenic emissions on the molecular composition of urban organic aerosols: An ultrahigh resolution mass spectrometry study, Atmos. Environ. 89 (**2014**) 525-532.

[85] Anonymous SANCO/10684/2009, Method validation and quality control procedures for pesticide residues analysis in food and feed. http://ec.europa.eu/food/plant/protection/resources/qualcontrol_en. pdf (accessed 01.10.13) (**2010**) .

[86] EUROPEAN COMMISSION, GUIDELINES FOR THE IMPLEMENTATION OF DECISION 2002/657/EC;SANCO/2004/2726-rev 4-December 2008, (**2008**) .

[87] D. Badocco, I. Lavagnini, A. Mondin, P. Pastore, Estimation of the uncertainty of the quantification limit, Spectrochimica Acta Part B: Atomic Spectroscopy 96 (**2014**) 8-11.

[88] D. Montville, E. Voigtman, Statistical properties of limit of detection test statistics, Talanta 59 (**2003**) 461-476.

[89] Gambaro A., Barbaro E., Zangrando R., Barbante C., Simultaneous quantification of microcystins and nodularin in aerosol samples using high-performance liquid chromatography/negative electrospray ionization tandem mass spectrometry, Rapid Commun. Mass Spectrom. 26 (**2012**) 1497-1506.

[90] Merel S., LeBot B., Clément M., Seux R., Thomas O., Ms identification of microcystin-LR chlorination by-products, Chemosphere 74 (**2009**) 832-839.

[91] Yuan M., Namikoshi M., Otsuki A., Watanabe M.F., Rinehart K.L., Electrospray ionization mass spectrometric analysis of microcystins, cyclic heptapeptide hepatotoxins: modulation of charge states and [M + H]+ to [M + Na]+ ratio, J. Am. Soc. Mass Spectrom. 10 (**1999**) 1138-1151.

[92] Diehnelt C.W., Peterman S.M., Budde W.L., Liquid chromatography-tandem mass spectrometry and accurate m/z measurements of cyclic peptide cyanobacteria toxins, TrAC Trends Anal. Chem. 24 (**2005**) 704-733.

[93] Welker M., Brunke M., Preussel K., Lippert I., von Döhren H., Diversity and distribution of Microcystis (cyanobacteria) oligopeptide chemotypes from natural communities studies by single-colony mass spectrometry, Microbiology 150 (**2004**) 1785-1796.

[94] Cortés-Francisco N., Flores C., Moyano E., Caixach J., Accurate mass measurements and ultrahigh-resolution: Evaluation of different mass spectrometers for daily routine analysis of small molecules in negative electrospray ionization mode, Anal. Bioanal. Chem. 400 (**2011**) 3595-3606.

[95] Namikoshi M., Rinehart K.L., Sakai R., Stotts R.R., Dahlem A.M., Beasley V.R., Carmichael W.W., Evans W.R., Identification of 12 hepatotoxins from a homer lake bloom of the cyanobacteria Microcystis aeruginosa, Microcystis viridis, and Microcystis wesenbergii: Nine new microcystins, J. Org. Chem. 57 (**1992**) 866-872.

[96] Sano T., Takagi H., Nishikawa M., Kaya K., NIES certified reference material for microcystins, hepatotoxic cyclic peptide toxins from cyanobacterial blooms in eutrophic water bodies, Anal. Bioanal. Chem. 391 (**2008**) 2005-2010.

[97] Mayumi T., Kato H., Imanishi S., Kawasaki Y., Hasegawa M., Harada K.-I., Structural characterization of microcystins by LC/MS/MS under ion trap conditions, J. Antibiot. 59 (**2006**) 710-719.

[98] Welker M., Maršálek B., Šejnohová L., von Döhren H., Detection and identification of oligopeptides in Microcystis (cyanobacteria) colonies: Toward an understanding of metabolic diversity, Peptides 27 (**2006**) 2090-2103.

[99] Falick A.M., Hines W.M., Medzihradszky K.F., Baldwin M.A., Gibson B.W., Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry, J. Am. Soc. Mass Spectrom. 4 (**1993**) 882-893.

[100] Zurawell R.W., Chen H., Burke J.M., Prepas E.E., Hepatotoxic cyanobacteria: A review of the biological importance of microcystins in freshwater environments, J. Toxicol. Environ. Health Part B 8 (**2005**) 1-37.

[101] Erhard M., Von Döhren H., Jungblut P.R., Rapid identification of the new anabaenopeptin G from Planktothrix agardhii HUB 011 using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 13 (**1999**) 337-343.

[102] Morrison L.F., Parkin G., Codd G.A., Optimization of anabaenopeptin extraction from cyanobacteria and the effect of methanol on laboratory manipulation, Peptides 27 (**2006**) 10-17.

[103] Rouhiainen L., Jokela J., Fewer D.P., Urmann M., Sivonen K., Two Alternative Starter Modules for the Non-Ribosomal Biosynthesis of Specific Anabaenopeptin Variants in Anabaena (Cyanobacteria), Chem. Biol. 17 (**2010**) 265-273.

[104] Ferranti P., Nasi A., Bruno M., Basile A., Serpe L., Gallo P., A peptidomic approach for monitoring and characterising peptide cyanotoxins produced in Italian lakes by matrix-assisted laser desorption/ionisation and quadrupole time-of-flight mass spectrometry, Rapid Commun. Mass Spectrom. 25 (**2011**) 1173-1183.

[105] Zafrir-Ilan E., Carmeli S., Eight novel serine proteases inhibitors from a water bloom of the cyanobacterium Microcystis sp. Tetrahedron 66 (**2010**) 9194-9202.

[106] J.C. Reepmeyer, D.A. d'Avignon, Structure elucidation of thioketone analogues of sildenafil detected as adulterants in herbal aphrodisiacs, J. Pharm. Biomed. Anal. 49 (**2009**) 145-150.

[107] M. Masiol, G. Formenton, A. Pasqualetto, B. Pavoni, Seasonal trends and spatial variations of PM10-bounded polycyclic aromatic hydrocarbons in Veneto Region, Northeast Italy, Atmos. Environ. 79 (**2013**) 811-821.

[108] M. Stracquadanio, G. Apollo, C. Trombini, A study of PM2.5 and PM2.5-associated polycyclic aromatic hydrocarbons at an urban site in the Po Valley (Bologna, Italy), Water Air Soil Pollut. 179 (**2007**) 227-237.

[109] I. Kourtchev, S.J. Fuller, C. Giorio, R.M. Healy, E. Wilson, I. O'Connor, J.C. Wenger, M. McLeod, J. Aalto, T.M. Ruuskanen, W. Maenhaut, R. Jones, D.S. Venables, J.R. Sodeau, M. Kulmala, M. Kalberer, Molecular composition of biogenic secondary organic aerosols using ultrahigh-resolution mass spectrometry: Comparing laboratory and field studies, Atmos. Chem. Phys. 14 (**2014**) 2155-2167.

[110] I. Kourtchev, J.-. Doussin, C. Giorio, B. Mahon, E.M. Wilson, N. Maurin, E. Pangui, D.S. Venables, J.C. Wenger, M. Kalberer, Molecular composition of fresh and aged secondary organic aerosol

from a mixture of biogenic volatile compounds: A high-resolution mass spectrometry study, Atmos. Chem. Phys. 15 (**2015**) 5683-5695.

13 Acknowledgment

First of all, I would like to thank my PhD supervisor Sara Bogialli, who gave me the opportunity to carry out my research and also for his support and guidance throughout.

I would like to thank all the members of the analytical chemistry group of the department of chemical Sciences of University of Padua: Paolo Pastore, Andrea Tapparo, Valerio Di Marco, Denis Badocco, Gabriella Favaro, Andrea Lentola Andrea Mondin, Maria Zulpo, Adriana Gatto and Iole Di Gangi; wonderful people that supported me scientifically and socially.

A special acknowledgement goes to Dr. Markus Kalberer who hosted me in his research group, giving me the formative and enjoyable opportunity to carry out some of my PhD program at University of Cambridge.

I would like to thanks also the mass spectrometry division of Kalberer research group for their support and helpful discussion during my research, namely, Chiara Giorio, Ivan Kourtchev and Brendan Mahon. I am grateful to Arthur Zielinski for the assistance with the Mathematica.

I would like to express my gratitude to all the wonderful people I have met during these three years around the world, all the teammates, all my friends and my flat mates.

I owe my deepest gratitude to my family who supported me in every moment of my education.

Finally, the most important acknowledgment goes to Angela, THANKS!