Università degli Studi del Piemonte Orientale "Amedeo Avogadro"

Dipartimento di Scienze e Innovazione Tecnologica
Dottorato di Ricerca in Chemistry&Biology
Curriculum: Energy, Environmental and Food Sciences
XXX ciclo A.A. 2016-2017

Target and non-target LC-MS/MS analysis of cosmetic, food and environmental samples

SSD: CHIM/01

Bianca Bolfi

Supervisor: Prof. Emilio Marengo
Co-Supervisor: Dr. Fabio Gosetti

PhD program co-ordinator: Prof. Luigi Panza

Università degli Studi del Piemonte Orientale "Amedeo Avogadro"

Dipartimento di Scienze e Innovazione Tecnologica
Dottorato di Ricerca in Chemistry&Biology
Curriculum: Energy, Environmental and Food Sciences
XXX ciclo A.A. 2016-2017

Target and non-target LC-MS/MS analysis of cosmetic, food and environmental samples

SSD: CHIM/01

Bianca Bolfi

Supervisor: Prof. Emilio Marengo

Co-Supervisor: Dr. Fabio Gosetti

PhD program co-ordinator: Prof. Luigi Panza

To my mother,
for being my strength

Contents

Chapter 4

Chapter 5

Chapter 1

Introduction

1.1 Purpose of chemical analyses and public safety

Analysis of complex matrices such as cosmetics, food and environmental samples still represents a difficult task since they are constituted by a mixture of compounds present in a wide range of concentration and characterized by different physical-chemical properties. Nowadays, most of the analytical methods are developed in order to be selective and sensitive for a specific class of compounds or a certain group of analytes, previously selected. Such *target approach* provides analytes detection and quantification through the use of selective analytical techniques, developed using reference materials (analytical standards). Method development, in fact, consists of different steps, including the sample preparation, matrix separation, instrumental analysis and data interpretation; each of them should be highly selective towards the target analytes. Once the method is developed, its parameters have to be optimized in order to provide the highest sensitivity, selectivity and specificity. For this aim, *a priori* knowledge of the investigated analytes is required.

Target approach is employed whenever the detection and quantification of particular analytes is required, even if the investigated concentrations are much less abundant than those of other components (Fig. 1.1).

The growing awareness that the information related to the molecules considered through target approach are not sufficient to provide a detailed picture of the

sample, i.e. to fully describe the complexity of heterogeneous matrices (such as cosmetic, food and environmental ones), led to the development of the *non-target approach*.

This enables to potentially identify all the molecules present in the sample, thus it considers all the substances, like for example degradation products and metabolites of analytes, that would have been excluded by the analyte-specific methods.

Non-target approach is useful, for example, in food safety [1, 2] and in the determination of potential hazardous compounds in environmental samples [1, 3-5]. Conversely, normative are focused on the monitoring of specific analytes, rigidly regulating their allowed concentrations in matrices, but they do not pose any attention on possible transformations that such analytes could undergo, caused for example by the effects of sunlight, temperatures, changes of pH or simply interactions with other compounds present within the same matrix. In particular, the term "*emerging substances*" indicates all the molecules that have been detected in a matrix, but that are not subject of restrictions since their fate and toxicity are not well understood.

Thanks to the modern techniques of separation and identification, specially to liquid chromatography coupled with high-resolution mass spectrometry, it is possible to investigate the state of conservation and pollution of complex matrices. In particular, the advent of high-resolution mass spectrometry (HR-MS) led to significant improvements of this approach. In fact, it enables the separations of analytes characterized by similar masses and provides accurate mass information, mandatory for the attribution of unequivocal molecular formula to unknown compounds. In order to achieve even better results, a good separation of the chemicals is necessary: liquid chromatography (LC) is the best candidate to this aim as it can reduce the chemical complexity, separate isomers and provide additional information (such as retention time), useful for the evaluation of the structure of unknowns [6]. In addition, unlike other separative techniques, it generally does not require any sample derivatization step prior analysis. Hence, the

hyphenation of LC to MS (LC/MS) represents the most suitable technique to perform *non-target* analyses. Such techniques will be discussed in Paragraphs 1.3 and 1.4.

As the most important aim of the non-target approach is the identification of as many compounds as possible, both sample preparation and the LC-MS method to be developed, must be suitable for different classes of compounds, hence they must be as much as unselective as possible. A great number of information will be available and it will have to be mined by means of powerful software, able to extract in a semi-automated mode the significant *m/z* signals to whom the operator will firstly attribute a molecular formula and then a chemical structure. In particular, the molecules that are identified through a search through databases, spectral matching and data available in the literature are defined as "*known unknowns*", while those for which a meticulous interpretation of the acquired data is needed to elucidate their structures are defined as "*unknown unknowns*".

Data processing, in fact, is the most time consuming step in *non-target* workflow since the number of signals is driven by matrix complexity and the number of compounds present therein. Conversely, the data quality is driven by the way of data collecting as it is mainly influenced both by the goodness of the chromatographic separation and by the mass detection.

After the attribution of a putative structure, it is always preferable to confirm the hypothesized structure with a reference standard (if available) or use additional information acquired by other analytical techniques able to elucidate molecular structures. Nuclear Magnetic Resonance (NMR) is one of the recommended techniques, even if it needs high concentration of the examined compound and total absence of other analytes (Fig. 1.1).

Non-target approach has been exploited in this thesis in order to investigate cosmetics (Chapter 3), food (Chapter 4), and environmental samples (Chapter 5), aiming to find out possible degradation products and other molecules that could be present in the investigated samples.

Lastly, *suspect-screening* approach can be described as an intermediate between target and non-target approaches. It consists, in fact, in re-interviewing previously acquired data, seeking compounds whose molecular structures are known, but that were not included in the target list when the analytical method was developed and validated. For example, it represents a useful approach when it is needed to point out substances that were not considered as hazardous when the analyses were carried out, but that afterwards were recognized as unsafe. New contaminants calculated *m/z* are extracted from the acquired data and their MS/MS spectra are compared with the spectra of the standards, if available. It is possible to build lists of relevant *m/z* on the basis of the literature dealing with the matrices that are intended to be investigated and on the experience of the analysts. The dedicated software (e.g. MasterView[TM], Sciex; ChromaLynx XS[TM], Waters; Sieve[TM], Thermo Fisher Scientific) automatically extracts the database calculated signals from the acquired chromatographic run and indicates with different colors (green, yellow and red) the positive, the tentative and the negative matching evaluating error percentage in mDa [7].

In this way, suspect and non-target screening approach can be performed at any time, without the need of the re-acquisition of the sample [7] (Fig. 1.1).

In this thesis, the *suspect-screening* approach was applied to environmental samples (Chapter 5) to evaluate the presence of previously identified degradation products of a fungicide used to protect rice crops.
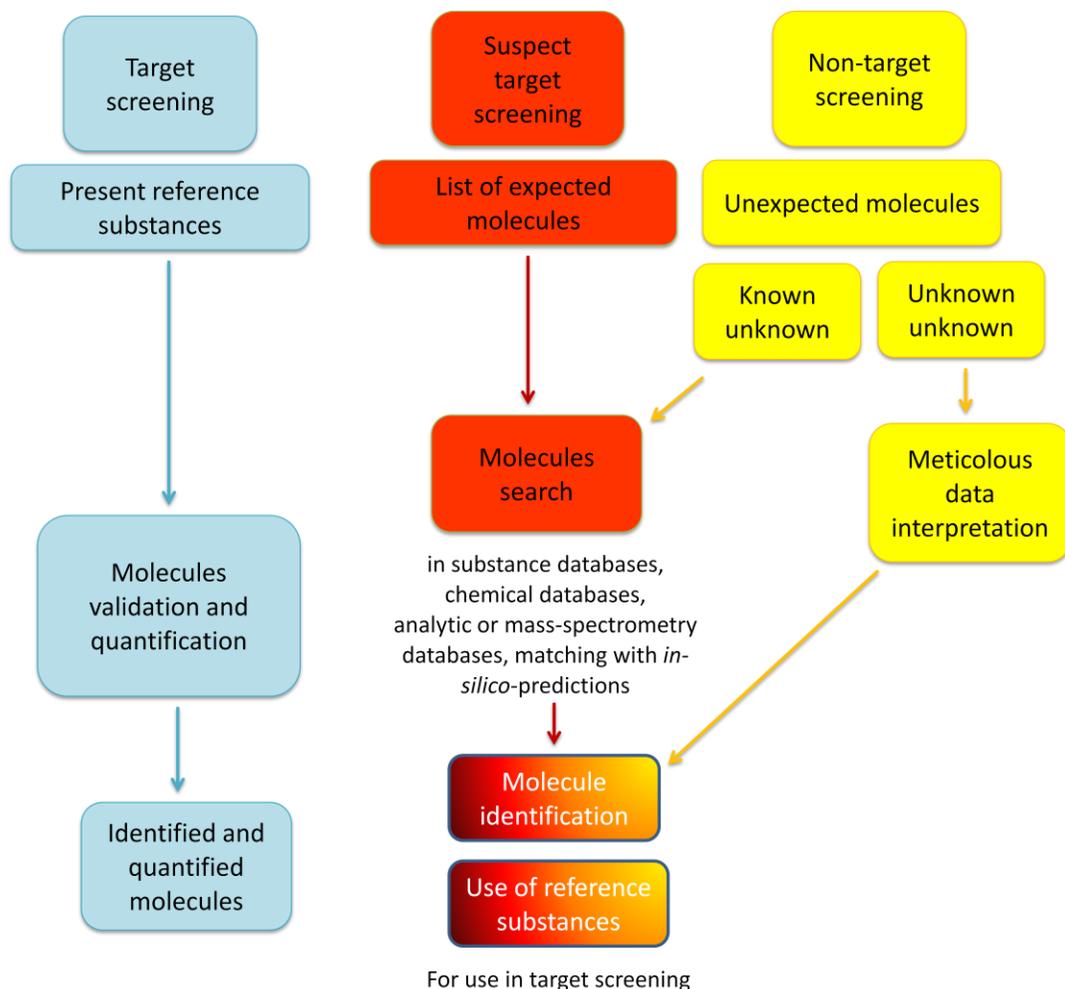
Figure 1.1: general scheme of target, suspect and non-target screening.

Target, non-target and suspect target screening can offer complementary information on the same considered sample, thus providing the potential monitoring of all the compounds present therein. This is extremely relevant to guarantee public health.

Monitoring and ensuring safety is one of the most important aim in research. Nowadays, liquid chromatography coupled with mass spectrometry is the golden standard for the investigation of organic contaminants in the majority of the analyte/sample matrix combinations occurring in different kinds of samples. To

ensure safety, in fact, it is of paramount importance to monitor samples considering the possible presence of both known contaminants recognized as hazardous and regulated by Normative, but also that of unknown, that could be due to *i*) impurities present in commercial formulations, generated by synthesis processes, *ii*) transformation products due to storage conditions or application of the product, *iii*) transformation products due to environmental conditions. In particular, in this thesis target and non-target LC-MS/MS approach was performed on cosmetic (Chapter 3), food (Chapter 4) and environmental (Chapter 5) samples.

Mass spectrometry finds application in cosmetic science not only for the analysis of the ingredients used in cosmetic products, but also for those related to biological samples, typically skin [8]. Cosmetic products are complex formulations designed to preserve healthy skin or to decorate it. Nowadays men use cosmetics as much as women do, while cosmetic uses is increasing also in babies and children. The European Union defines cosmetic products as any substance or mixture intended to be placed in contact with different external parts of the human body or with the teeth and the mucous membranes of the oral cavity with a view exclusively or mainly to cleaning and perfuming them, changing their appearance, correcting body odor, protecting them, and keeping them in good condition [9]. Basing on such definition, personal care products (i.e. cream, soap, perfume, lotion, deodorant), beauty products (eye shadow, mascara, lipstick, lip gloss, nail polish), oral products (mouthwash, toothpaste), and temporary decoration are considered as cosmetics. The ingredients present in cosmetics are directly related to their use as pigments, fragrances, absorbers, antioxidants, UV filters. Some hazardous substances could be present in the final formulation because of impurities in raw materials, side synthesis reactions, ingredient degradation products or even illegal addition. Thus, cosmetic must be object of safety evaluation, considering the current Regulations. Scientific Committee on Consumer Products (SCCP) periodically review the restricted substances whose maximum concentrations in different formulations are listed in Annex III-VI of Council Directive 76/768/EEC,

while the non-allowed toxic compounds are listed in Annex II of Council Directive 76/768/EEC [10]. Even if more than 1200 nonallowed compounds for cosmetic formulations, including antibiotics, hormones and glucocorticoids, same producers intentionally add them to their products, posing a serious risk to consumers [11, 12]. Many metals can be present in cosmetics such as make-up powders [13], lipsticks [14], eye shadows [15] and care creams [16]. Some metals (i.e. nickel, cobalt, chromium and palladium) can lead to contact dermatitis [17] as they can accumulate in skin stratum corneum, while others (i.e. mercury, lead, cadmium and aluminum) may reach blood streams passing skin layers and be transported to organs where they can be accumulated, leading to toxic effects [18]. Since cosmetics are freely applied on skin without restrictions and may reach organs, it is of paramount importance to monitor the composition of the ingredients used in commercial formulation, focusing on their chemical structures and their possible transformations that can occur due to the interaction with skin or simply by modification due to external agents (i.e. sunlight).

In this framework, it is clear how mass spectrometry represents an indispensable analytical tool to ensure safety for consumers, not only thanks to its sensitivity and specificity towards target compounds, but also thanks to the opportunity of investigating the effects of external agents on intact cosmetics formulation, adopting non-target and suspect target screening.

In this thesis, the nonpermanent tattoo cosmetic matrix was considered (Paragraph 3.1). In particular, the Sulforhodamine B organic dye was investigated, focusing on the identification of its possible degradation products when applied on skin and exposed to sunlight irradiation (Chapter 3).

Similarly, safety and quality of food products are of growing interest for consumers, governments and producers. Food manufacturers can prevent, reduce, control and eliminate possible food borne hazards adopting the good manufacturing practices (GMPs). Furthermore, critical points in food production are highlighted by the Hazard Analysis and Critical Control Point (HACCP) system, aiming to

guarantee prevention, mitigation and control of food contamination. In addition, analytical information focused on recognized and new identified contaminants is of paramount importance, since food could be contaminated by different sources, such as environment, industrial practices, agrochemical residues or toxins [19]. Moreover, food frauds and adulteration are still critical [20]. In recent years, many publicized incidents worried consumers, such as the bovine spongiform encephalopathy in beef, benzene in carbonated drinks in UK, dioxins in pork and milk products from Belgium and pesticides in soft drinks in India. To guarantee safety, many Governments re-evaluated their attention to food safety, increasing the number of food safety-related legislation [21]. It has to be underlined that, even if there is no world harmonization in Food Safety legislation, FAO and WHO established the *Codex Alimentarius Commission*, whose aim is to develop science and establish standards for international trade to be considered by Countries in their legislation [22]. The European Union, that considers food safety as one of its priority, published a White Paper on Food Safety to ensure safe products, evaluating their production, processing, storage, transportation and sale [23].

The Food Safety legislative framework establishes sampling and monitoring plans, the maximum residue limits (MRLs) for contaminants and residues, the minimum required performance limits (MRPLs) for the adopted procedures to detect the forbidden substances, and the performance characteristics of the analytical methods [21]. Thus, development, optimization and validation of analytical methods are the fundamental steps in testing food quality. The main topics related to food analysis are related to the composition and properties of food products, beverages, flavors and physico-chemical changes that they can undergo due to preparations, processing and storage conditions. To ensure the safety of food it is necessary to monitor not only the sensory and nutrition properties or the assess the presence of contaminants, but also evaluate the stability of the products, since it can limit their shelf life.

Considering this background and need, LC-MS finds application in food safety testing. In particular, it is well suited for the analysis of food contaminants, as it provides a great number of information about the heterogeneous mixture the components constituting the food product. The application of LC-MS in food safety allows a more comprehensive evaluation not only of the presence of target contaminants and residues at trace levels, but also that related to unexpected or unknown compounds that might contaminate the product.

In particular, in this thesis, the quality of extra virgin olive oil (paragraph 4.1) was evaluated considering nine target compounds concentration trends and investigating the presence of possible contaminants [Chapter 4] when stored in different packaging materials, in order to evaluate its conservation in time.

Not only cosmetics and food are exposed to contamination, but also environment, in particular water (Paragraph 5.1). Both surface water and groundwater are polluted by contaminants due to domestic, agricultural, industrial, pharmaceutical and natural sources. Water contaminants are characterized by different chemical properties, except for their solubility to certain values in water. There are three main different groups in which it is possible to divide water organic contaminants: pharmaceuticals, microcystins and pesticides. Even if a high number of pharmaceuticals are detected in screening studies addressed to the evaluation of water quality [24, 25], pharmaceuticals monitoring is not routinely. Microcystins are one of the most frequently cyanotoxins found in water. They are secondary metabolites of cyanobacteria, one of the most important algal classes in surface water. Microcystins are molecules characterized by high stability and persistence, due to their cyclic structure. They can cause severe poisoning not only to animals, but also to humans if their concentration in drinking water is too high [26]. Nowadays more than 60 microcystins are characterized, all of which are hepatotoxic cyclic peptides [27]. The last group of human waste pollutants is represented by pesticides, widely used since the middle of 1900. Actually, there are more than 860 active principles used in pesticides formulations, belonging to more

than 100 different classes, e.g. pyrethroid, triazines, carbamates, bonzoylureas, sulfonylureas [28]. Their different physical-chemical properties (i.e. some of them are characterized by the presence of halogens, nitrogen, sulfur, phosphorous) are important when analytical methods have to be developed for their detection, in particular they could address the choice of the best instrument that has to be adopted (i.e. to choose between gas chromatography or liquid chromatography). In all the developed Countries, different Regulations enforce MRLs of the different contaminants. In past years, contaminant monitoring was mainly performed using gas chromatography coupled with mass spectrometry (GC-MS and GC-MS/MS), requiring a multistep sample preparation. In recent years, liquid chromatography tandem mass spectrometry replaced GC-MS/MS, becoming a routine analysis technique, enabling the monitoring of a wide range of polar and nonvolatile substances that would have been ignored by GC-MS analysis and reducing the required sample preparation. In fact, the use of LC-MS/MS does not need any preconcentration step, ensures better sensitivity, leading the governmental organs to adopt same or lower MRL of contaminants.

Chapter 5 deals with LC-MS/MS analysis of paddy water samples, in order to assess the overall impact of rice treatments on water quality. The LC-MS/MS technique enables the performing of both target analysis focused on the identification of three pesticides used in rice cultivation and non-target and suspect target screening to consider paddy water in all its complexity.

1.2 Sample pretreatment

The ideal sample is representative, homogeneous, without interferences in order to fit the requisites for injection into the LC column, to be compatible with the designed analytical methods and to guarantee the quantitative recovery of the target analytes. Moving from the ideal sample to the real one means adding many variables. Samples are the result of complex interacting chemical classes,

characterized by different structures, polarities, masses, acid/basic properties, concentrations, etc.. As each matrix has its own characteristic, it has to be considered in all its complexity.

The sample pretreatment, designed for target analyses must be highly specific for the selected compounds as it aims at their detection and quantification. To investigate analytes present at very low concentration (i.e. sub trace levels), it may be useful to eliminate compounds that might interfere with the LC-MS analysis. Sample pretreatment represents often a critical step, as it can affect LOD, LOQ, recovery, interferences, selectivity and specificity.

Conversely, the extraction method that must be considered in non-target sample pretreatment has to be highly unselective in order to evaluate the presence of all the possible compounds. For these reasons all the manipulations have to be reduced, preventing the possible losses of the unknown species. The reduced sample manipulation guarantees the preservation of all the components present in the sample, but some species could interfere with the analytes during the chromatographic analyses. In fact, when considering LC-MS quantification, the matrix effect has to be evaluated because it can enhance or suppress the analyte signal. Co-elution could take place and modify the ionization in the source. Furthermore, some components could alter the signal because they are in competition with the analytes in the formation of the charged drop in ESI mode (Paragraph 1.4.2) [6]. What researchers can do, is "dilute and shoot". In this way, the matrix is diluted reducing the ESI suppression, but it preserves its original composition that is interesting for non-target detection. This also represent an economic advantage because this safeguards columns duration. Another important plus is represented by the conservation of the sample structure, since it does not modify the existing relationships between the chemical species that constitute it. In this way, it is possible to obtain an effective fingerprint of the sample with regard to all of its constituents. The dilution factor has to be carefully evaluated, because of the complexity of the matrix. As described above, real samples are the result of

an heterogeneous mixture of chemicals present at different concentration levels. Diluting too much could lead to the non-identification of some compounds, while the non-dilution could cause signal saturation. It could appear as a nonsense, but the best way of sample preparing is no sample preparing at all.

On the contrary, if the sample is available in considerable amount, the best way to handle it, is portioning and addressing the portions to different pre-treatments, in order to highlight the different classes of analytes present, making them more evident using different chromatographic conditions. All those approaches are addressed to the fulfillment of a sample that is exhaustive and does not damage the instrument. If the sample is particularly rich in interferent compounds, a cleanup is necessary. To this aim, one of the most pertinent pretreatment techniques is SPE (Solid Phase Extraction), evaluating the optimization of the sorbent, washing agents and eluents.

Furthermore, it is advisable to use an internal standard (IS) in order to check the possible instrumental response changes. The chosen IS should cover the entire retention time (RT) and mass range designed by the LC-MS method in both ionization modes. It is useful to fortify samples with an internal standard (using a substance that must not be naturally present in samples to be analyzed), adding it in a known amount. Variations in measurement performance over longer time spans and batches can be found out using the IS [29]. Through this, it is possible to check the suitability and stability of the analytical method: in fact, this could highlight possible variations in retention times as well as intensities of the signals. If this happens, it is possible to align the whole analysis to the original one. It can be seen as a way to check the sensitivity of the method as well as the quality of the response along time (to find out effects from, for example, technical changes such as replacement of the ESI needle, the detector and its voltage, or different operational status of the system concerning variations of the involved pressures, temperatures etc.) [30].

1.3 Liquid Chromatography

Liquid Chromatography (LC) is one of the most adopted quantitative techniques in the analytical laboratories. Many technological efforts have been addressed to the improvement of this technique in terms of speed of the chromatographic separations and sensitivity. In particular, a reduction of the particles diameter (d_p) packed into the chromatographic columns provided better efficiency, mass transfer and speed. On the basis of the Plate Model, the efficiency N is defined as the number of theoretical plates, calculated as the ratio between the column length L, and the height equivalent to a theoretical plate, H:

$$N = \frac{L}{H}$$

A column is much more efficient as the number of theoretical plates is higher, but for the same length L, H must be the smallest possible. In particular, H is mathematically expressed through the Van Deemter equation:

$$H = A + \frac{B}{u} + Cu = 2\lambda d_p + \frac{2\gamma D_m}{u} + \frac{f(K)d_p^2 u}{D_m}$$

where:

u is the linear flow rate or linear velocity;

λ is the packaging constant;

γ is the obstruction factor for diffusion in column;

D_m is the analyte diffusion coefficient in mobile phase;

f(K) is a function of analyte retention factor.

The *A* term (Eddy diffusion) is referred to the multiple flow paths: during elution in fact, different molecules of the same analyte can follow different paths within the packed column, so they can be eluted at different times, leading to bandwidth widening and a decrease of efficiency. The availability of smaller diameter particles favors a homogeneous packing, that is reflected on the homogeneity of the

paths that molecules can follow, thus lowering the *A* coefficient. The longitudinal diffusion (along the axis of the column) that occurs passing from areas characterized by high concentrations to the low ones, is expressed by the *B* term. To reduce this factor it is necessary to increase the linear velocity, leading to the reduction of the time spent by the analytes in column. Term *C* is related to the resistance to the mass transfer in the mobile phase: it is the parameter that is mostly influenced by the particles size because of its direct correlation with the square of the particles diameter. Hence, it is possible to lower the *C* term by reducing the size of particles, thus decreasing the height of the theoretical plates (*H*) and increasing the efficiency (Fig. 1.2).

It is possible to determine the minimum of the Van Deemter curve corresponding to the best linear velocity conditions through the following differential calculation:

$$\frac{dH}{du} = 0,$$

valid for

$$u_{opt} = \sqrt{\frac{B}{C}}$$
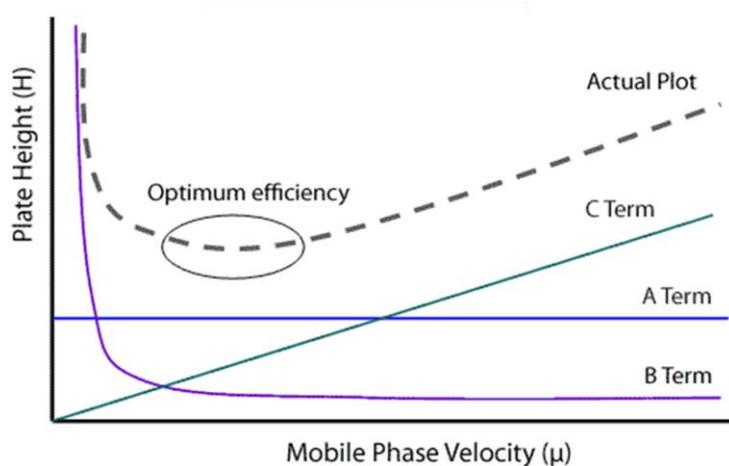
Figure 1.2: relationship between mobile phase velocity and efficiency expressed as the height of theoretical plates.

Sub-2 µm particles [31] were first used by Waters Corp. (Milford, MA, USA) in 2004 for Ultra-Performance Liquid Chromatography (UPLC) system, then commercialized also by other vendors as "ultra high performance liquid chromatography" (UHPLC), widely accepted as a "high throughput" equipment [32]. As reported in Figure 1.3, a dramatic increase of efficiency is provided by the use of sub-2 µm particles when increasing flow-rate. The overall effect is the analysis time reduction and efficiency improvement in UHPLC respect to HPLC.
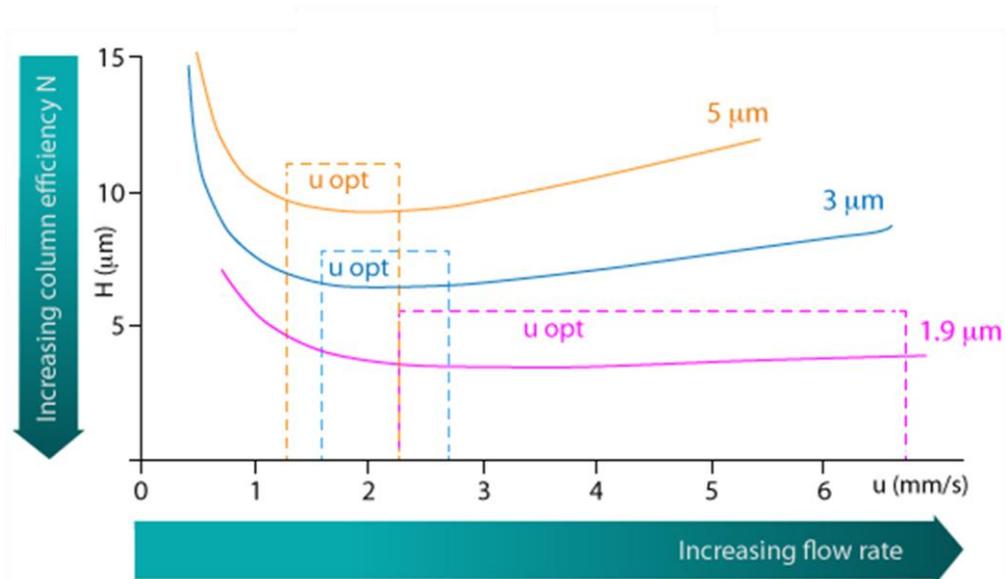
Figure 1.3: efficiency variation as a function of flow rate.

However, UHPLC requires high inlet pressure to induce the mobile phase to flow into the column, since it is directly related to the square of the d_p (Darcy's law) and to the cube of d_p when u_{opt} is adopted [33]. Moreover, according to Darcy's law, the flow rate is directly proportional to backpressures [33]:

$$\Delta P = \frac{F\eta L}{K^0 \pi r^2 d_p^2}$$

ΔP = backpressure; K^0 = specific permeability;

F = flow; r = column radius

η = viscosity; d_p = particles diameters

L = column length;

Technological efforts have then been addressed to overcome this problem: *core-shell* stationary phases were developed and commercialized in 2001 by Agilent Technologies, the *Poroshell 300*, characterized by 5 μm particle size. Six years later, particle size was reduced to 2.7   μm in Halo (Advanced Materials Technology), Ascentis Express (Sigma-Aldrich) Kinetex and Aeris (Phenomenex) columns [34]: their success led to market competition and consequent commercialization of Accucore (Thermo Scientific), Poroshell 120 (Agilent), Cortecs (Waters) etc.

The new particles are constituted by a solid silica core (0.9-3.7 μm) and a thick shell (0.4-1.3 μm) of porous silica material. The core-shell technology provided an increased loading capacity with respect to the previous one, the fully porous silica [35] (Fig. 1.4).
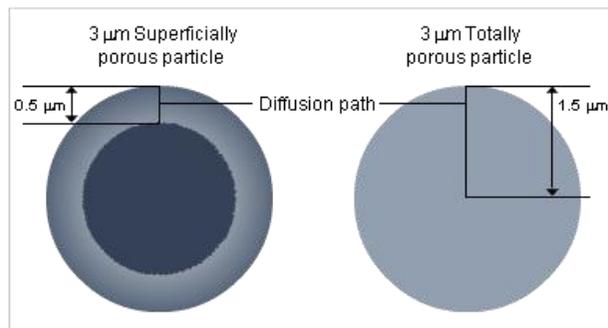
Figure 1.4: superficially porous and totally porous particles.

Furthermore, they reduced not only the *C* and *B* terms in van Deemter equation, respectively related to the resistance to mass transfer and longitudinal diffusion, but also the *A* one, related to Eddy dispersion, since the homogeneous and dense packing is able to reduce the number of possible paths that analytes follow when eluting from the column [36-42]. In particular, the *B* term reduction is ascribable to the solid core constituting the 25% of each particle that inhibits the longitudinal diffusion [43].

For compounds characterized by low molecular weight, the minimum height of the theoretical plates values ranged from 2.6 to 4.3 μm [44]. Successively, further particles size reduction was achieved, providing 500.000 plates/m [45, 46]. In particular, the 1.6 μm core-shell particles demonstrated to be more permeable and stable at high pressures (1200 bar) [35]. Despite the advantages provided by the small particles diameters above described, 4-5 μm core-shell particles are still popular in routine laboratories that have to follow certified methods (for example pharmaceutical industries use Pharmacopeia methods). Such packing is highly permeable, resulting to be suitable for high resolution separations [35]. It can be concluded that columns with core-shell packing provided several improvements [47]:

- they can withstand high flow rates at low pressures, thanks to their improved permeability;
- the dense and homogeneous packing of the column reduces the ability of the analytes to follow different paths within the column itself;
- longitudinal diffusion is reduced by the solid core;
- the thick porous layer surrounding the core improves peaks shape since it reduces the band broadening due to diffusion and intra-particle paths;
- they avoid the formation of thermal gradients generated by the adopted high flow rates, since the solid core is highly thermally conductive. It results in

better and homogeneous analytes migration velocity across the column, leading to narrower peaks [48].

Retention factors and separations are influenced by analytes-stationary phase interactions which in turn are influenced by the particles diameters. In particular, given the design of the core-shell particles, it is possible to achieve better separations since selectivity and retention are improved by the low amount of porous material that surrounds the particles, making them less hydrophobic than the totally porous ones.

When analyzing complex matrices characterized by the presence of a very large number of compounds, the *peak capacity* has to be taken into consideration. It expresses the maximum number of peaks in a specific chromatographic space with unitary resolution for all the adjacent peaks. According to the given definition, it should be a constant parameter, thus not influenced by flow rate. A study conducted by Gritti et *al.* [49] compared Halo C18 and Kinetex C-18 (two core-shell packed columns) with BEH C18 (totally porous packing): at low flow rates, the columns did not show significant differences in their peak capacities, while increasing speeds, only the Kinetex column maintained unchanged its performances. Further studies conducted by the same authors [50, 51] compared peak capacities and efficiencies of monolithic, superficially porous (2.7 μm) and totally porous (1.7 μm) columns using fast gradient elution. The results demonstrated comparable peak capacities in the range of 27-30 for all the tested columns.

Temperature is another important parameter to consider in order to improve the chromatographic performance and to reduce the backpressures. In fact, setting the column temperature higher than the room one (temperature higher than 50-60 °C can affect columns stability) makes the mobile phase more fluid as it reduces its viscosity. On the basis of Darcy's law, the use of low viscosity mobile phase should be preferred in order to prevent high pressures. If we consider the reversed-

phase LC as an example, acetonitrile should be adopted as organic part in mobile phase instead of isopropanol, whose viscosity is much larger than the acetonitrile one (η_{iso}=2.43 mPa·s; η_{ACN}=0.39 mPa·s). Thanks to this, higher flow rates can be adopted without losing efficiency [52-54].

It is possible to resume all the effects that lead to peak broadening (*W*) as the sum of those due to column (w_c) and to all the processes that take place outside the column, defined as *extra column effects* (w_{ec}):

$$W^2 = w_c^2 + w_{ec}^2$$

Extra-column peak broadening has to be minimized to further improve the chromatographic separation of the analytes. In fact, loss of efficiencies could be caused by non-optimized LC-systems using core-shell particles [55]. In order to reduce the w_{ec}^2 term, for example, the pumping system, column and detectors should be as close as possible to each other, in order to reduce the length of the connection tubing. The use of small internal diameter tubing (i.e. 0.12 mm) should be adopted; the correct assembling of all the fittings has to be controlled carefully; detector cell and data collection rate have to be evaluated. In fact, the detector must be fast enough to guarantee the acquisition of an appropriate number of data points for each chromatographic peak.

The adoption of the described strategies improves also the peak shape and reduces the solvent consumption [47].

1.4 Mass spectrometry

Mass spectrometry is an analytical technique particularly suitable for the quantitative analysis of target compounds, for the identification of unknown molecules and their structure elucidation. It is possible to obtain all these information with a very low amount of sample (in some cases of the order of picograms), with very low analytes concentration in complex matrices (down to parts per trillion). It is commonly used in combination with separative techniques like liquid chromatography. The principle on which mass spectrometry is based is the possibility of separating ions depending on their mass-to-charge ratio, through the use of magnetic or electric fields or radio frequencies. The mass spectrum is the diagram in which the abundance of each ion is reported as a function of the mass-to-charge ratio. The mass spectra are different for every compound and they are also highly dependent on the method adopted for ion formation.

1.4.1 General description of the instrumental components

The mass spectrometers hyphenated to liquid chromatographic systems consists in (Fig.1.5):

- an inlet system;
- an ion source;
- a mass analyzer;
- a detector;
- a system for signal elaboration.
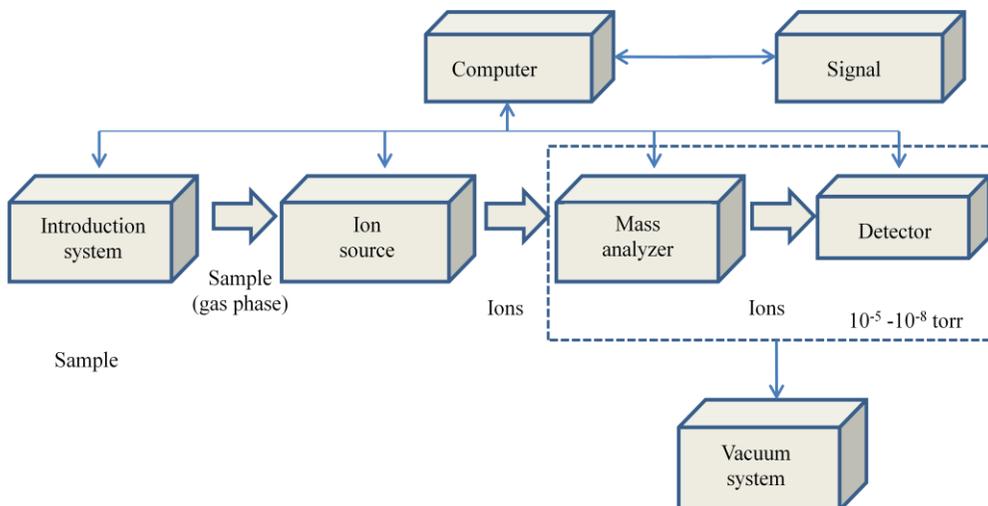
Figure 1.5: scheme of a LC/MS system.

The *inlet system* allows to introduce into the ion source the minimum amount of sample required for the analysis, transforming the molecules in gaseous ions. For this aim, different devices able to volatilize liquids are commercially available.

The *ion source* aims to convert the molecules present in the sample into ions. Through this step, the positively or negatively charged ions beam is accelerated to the mass analyzer by the electrical field.

The *mass analyzer* selects the ions, on the basis of their mass-to-charge ratio (*m/z*): *m* is the mass number of the ion expressed in atomic mass unit, while *z* is the number of charges present on the ion.

The beam of ions is converted into an electrical signal thanks to the *detector*: only through this step the signal can be elaborated, acquired and recorded in a personal computer and then it can be used by the analyst.

It is of great importance that all the parts of the instrument, except for the sample introducing system and the ion source, are kept under high vacuum (from 10^{-4} to 10^{-8} Torr): such a condition avoids collisions between electrons and ions and other gaseous molecules present within the system.

1.4.2 Gaseous ions formation: Electrospray ionization

In this thesis, the Electrospray ionization (ESI) (Fig. 1.6) technique was adopted: it is characterized by the small energy imparted to the molecules, which leads to a reduced fragmentation. Thus, the obtained mass spectra are characterized by the presence of quasi-molecular ion peak, while other peaks are rare or totally absent. ESI represents one of the most widely used techniques both in "omics" analysis (proteins, polypeptides of thousands of Da) and smaller molecules (polar molecules containing heteroatoms, organic acids and bases or salts). The ionization takes place in a source similar to that reported in Figure. 1.6, at atmospheric pressure.
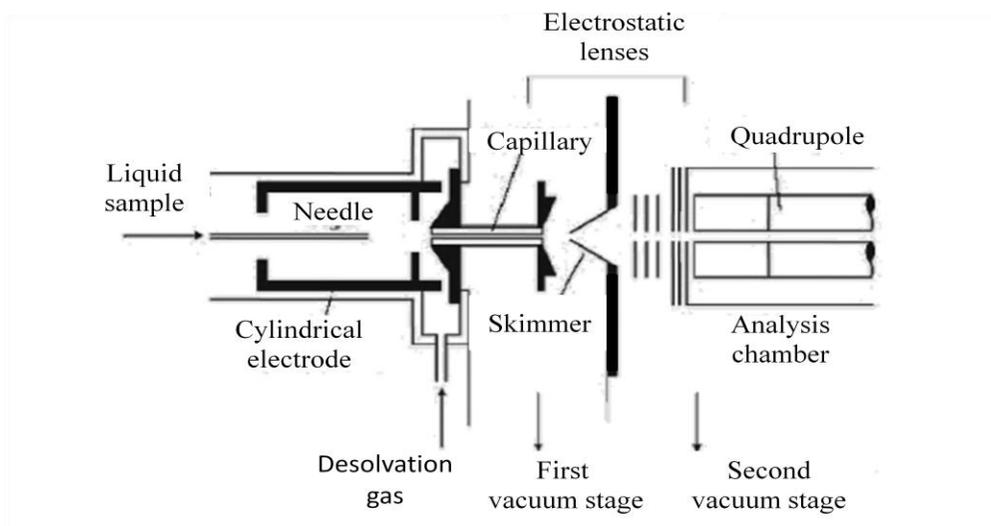
Figure 1.6: scheme of electrospray source.

The liquid sample is introduced through a stainless steel capillary, kept at 4000-5000 V with respect to the electrode that surrounds it. The liquid is directed from the ESI metallic tip in the *Taylor cone*. In fact, when a small amount of electrically conductive liquid is exposed to an electric field, the shape of the liquid surface begins to deform due to the loss of surface tension. The applied voltage increases the force that acts on the drop surface tension, generating a cone. The electric field

and the electrostatic repulsion between the charged molecules into the Taylor cone becomes too strong and leads to passing the surface tension limit, the *Rayleigh limit*. At this point, the Coulomb repulsion of the surface exceeds the liquid surface tension, inducing the liquid to exit from the cone as drops that are broken into smaller ones, generating the spray. The described process is reiterated until the ion desorbs from the surface of the droplet (ion evaporation model) [56, 57] or the solvent is completely removed (residual charge model) [58, 59]. The exact mechanism that regulates the ions formation, both in the case of evaporation of the ions or in the complete removal of the solvent, is still argument of debates. Anyway the result is that thanks to the voltage applied, the droplets can move from the tip towards the counterelectrode [60] (Fig. 1.7).
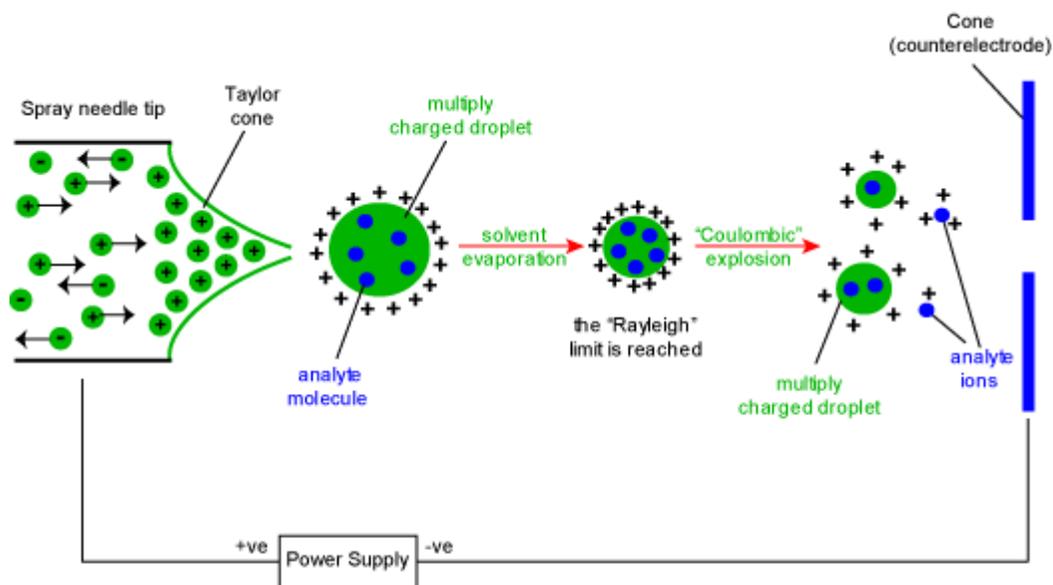
Figure 1.7: scheme of the "ion evaporation" mechanism.

Inert gases are continuously flowed in the MS entrance. The curtain gas (CUR) provides a clean and stable environment for the ions of the sample that enter into the mass spectrometer. It does not allow air or solvent to enter into the analyzer

region of the instrument, allowing the sample ions to enter into the vacuum chamber thanks to the effect of the electric fields, preventing possible contamination of the ion optics. Traces of moisture or solvent vapors negatively affect the analysis as they cause both a decrease in sensitivity, stability and resolution, as well as an increase in background noise.

The nebulizer gas (GS1) favors the formation of small droplets, that rapidly desolvate in the ion source. The heater gas (GS2) is used to improve the ionization efficiency, helping the spray droplets to evaporate. The temperature of GS2 and its flow are proportional to the flow of the incoming liquid and the percentage of the aqueous solvent contained therein. It is important to underline that too high temperatures can lead to an increased background noise, as they cause early vaporization of the solvent; the adoption of a too high gas flow can lead to an unstable signal.

Thermally labile and big biomolecules are fragmented in a minor extent in the ESI process, because of the low supplementary energy retained by the analyte during the ionization process. When the ions are multiple-charged, the corresponding m/z values are low enough to be detected by the analyzer.

The efficiency of the ESI technique depends on the number of the detected analytes compared to the nebulized ones and on the desolvation, ionization and transfer processes into the vacuum system. It can be deduced that the smallest are the particles, the highest charge density and the lowest desolvation energy they have.

1.4.3 Analyzers

The mass spectra can be considered as a representation of the ions generated in the ionization source, characterized by different m/z ratios: it is necessary to separate the generated ions, to determine the m/z values and then to measure the relative abundance of each group of different ions. Different devices, the analyzers, are able to separate the ions characterized by different $m/z;$ each analyzer is characterized

by its strengths and drawbacks. Among the most used analyzers there are: ion traps, quadrupoles, magnetic sectors, times of flight, orbitraps, FT-ICR and hybrid analyzers, i.e. combination of two or more of the previously mentioned ones. This section will be focused on the analyzers used for this thesis: a hybrid Q-Trap used in triple quadrupole mode and a hybrid QTOF (quadrupole-time of flight). The parameter usually employed to evaluate the analyzer analytical capabilities are their resolution and mass accuracy. The *resolution* (R) is referred to the capability of distinguishing two consecutive signals, or two *m/z* ratios of different ions; it is mathematically expressed as:

$$R = \frac{m}{\Delta m}$$

where *m* is the nominal mass of the first peak and Δm represents the difference (in Da) between two consecutive resolved peaks.

The level of information we can obtain from a mass spectrometer depends on its resolution. Low-resolution instruments provide only the nominal ion mass, while high-resolution instruments provide the exact mass of ions, which generally uniquely define the elementary composition of the corresponding ions.

The reason is because the isotopes masses are not exact integer, for example ^{14}N has a mass of 14.0031 Da, ^{16}O 15.9949 and ^{12}C 12.000 Da, hence CO and N_2 (both with nominal mass of 28 Da) are 27.9949 Da and 28.0062 Da, respectively, thus to separate CO^+ from N_2^+ a resolution of at least 2500 is required. To resolve peaks with lower *m/z* ratio differences, instruments with higher resolution are required. Another parameter that characterize the analyzers, is their mass accuracy, expressed as:

$$\Delta m_{accuracy} = m_{true} - m_{measured}$$

This parameter is deeply influenced by the quality of the calibration: it evaluates the deviation of the response of the instrument from a known monoisotopic calculated mass.

Accurate mass "dynamic calibration systems" is employed by the LC-TOF-MS, for example, to provide accuracies for fragment ions better than 3 ppm. To improve robustness in mass accuracy measurement along time, reference sprayers, such as lock-spray, are used. In fact, environmental factors can generate drifts in the mass measurements, leading to the instability of the calibration. For this reason, the daily calibration of the mass spectrometer is necessary [3, 61].

In this thesis a hybrid Q-Trap used in triple quadrupole mode and a hybrid QTOF (quadrupole-time of flight) were respectively used for target and non-target analyses of food, environmental and cosmetic samples.

1.4.3.1 Quadrupole

The quadrupole operates using a dynamic electric field that allows only the ions with a specific *m/z* ratio to be transmitted, hence to obtain a mass spectrum it is necessary to operate in scan mode so that different ions can be sequentially transmitted. The quadrupoles require a continuum ionic beam, resulting to be highly suitable for the hyphenation with an ESI source.

Quadrupoles mass analyzers are constituted by four metallic rods (generally in molybdenum), aligned along the *z* axis. The ideal analyzer should have hyperbolic bars, able to create a linear and symmetric field [62] (Fig. 1.8.A), but the majority of the commercial analyzers have cylindrical bars (Fig. 1.8.B).
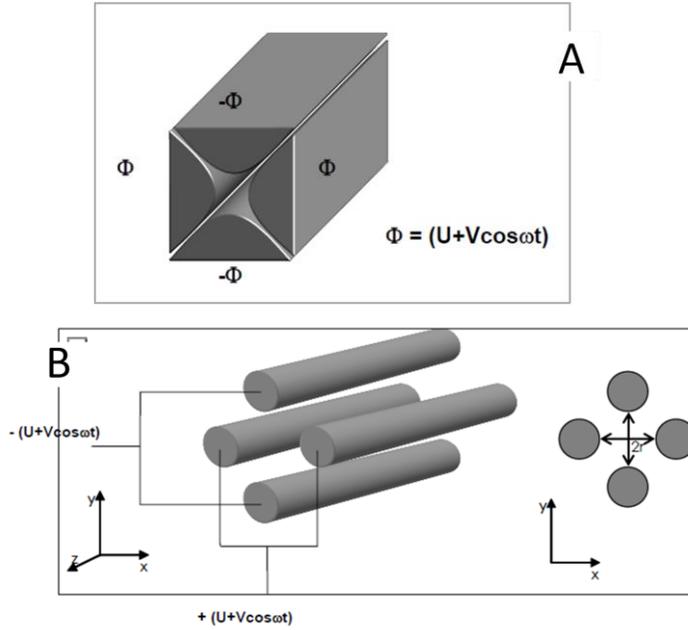
Figure 1.8: scheme of the ideal quadrupole analyzer characterized by hyperbolic bars (A), and of the real one (B), characterized by cylindrical bars.

A voltage is applied to each pair of electrically connected bars: U in the case of a continuous electric field (DC), and V in the case of an alternate RF (RF). The ions, coming from the source and entering the quadrupole, are affected by acceleration forces in the *x* and *y* directions induced by the combined electric fields. The motion of an ion through this field is described by the Mathieu equation that defines the stability and instability regions for the ions in an electric field generated by a quadrupole. Considering the motion of an ion characterized by *m* mass and *q* charge, the equation of motion on the three axis are described as follows:

$$\frac{d^2x}{dt^2} = -\left(\frac{e}{m}\right)\frac{[U+V cos(\omega t)]}{r_0^2}x,$$

$$\frac{d^2y}{dt^2} = \left(\frac{e}{m}\right)\frac{[U+V cos(\omega t)]}{r_0^2}y,$$

$$\frac{d^2z}{dt^2} = 0.$$

where:

e= charge of the electron ω= radiofrequency

m = mass V= radiofrequency voltage

r_0= distance between the opposite rods U= direct current voltage

These equation defines stability areas that are functions of U and V, for ions with different *m/z* ratio; the stability diagram (Fig. 1.9) reports on the *x*-axis the *q* value and on the *y*-axis the *a* value, calculated as:

$$q = \frac{4eV}{mr_o^2\omega^2}$$ $$a = \frac{8eU}{mr_o^2\omega^2}$$

Varying the arguments of the two equations, and keeping the U/V ratio constant, it is possible to stabilize and thus allow the crossing of the quadrupole of a specific *m/z* ion. Quadrupoles mass analyzers, in fact, can be considered as mass-selective filters, able to transmit only the selected ions. Given specific combinations of DC and RF voltage, quadrupoles are able to transmit only the ion characterized by a specific *m/z* ratio in a certain range, while other ions hit the rods and are filtered out. Performing a scan on the mass range, it is possible to achieve the separation and subsequent detection of the different masses.
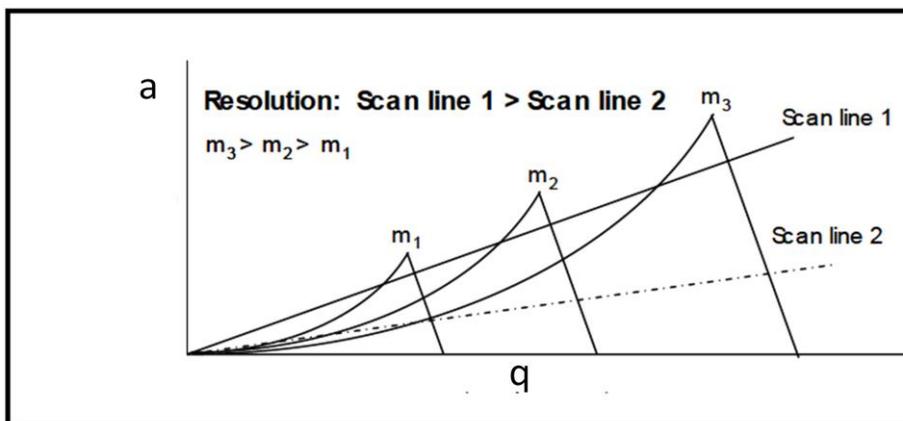
Figure 1.9: stability diagram for three different ions characterized by a different *m/z* ratio.

Three quadrupoles can be used in a linear series, constituting the scheme of the triple quadrupole mass spectrometer (QqQ), where the first (Q1) and the third (Q3) quadrupole act as mass filters, whereas the second one (Q2) acts as a collision cell (Fig. 1.10). In Q2, N_2 or Ar ($\sim 10^{-3}$ Torr) are employed to produce collisional dissociation of the ions selected in Q1. The generated fragments are addressed to Q3, where they can be scanned or filtered again. The described process allows the study of fragments, the *product ions,* that are essential in structure elucidation. For example, Q1 could be employed as a filter towards a specific ion with a known *m/z*, that is subsequently fragmented in Q2. Q3, working in scan mode in the *m/z* range, could provide the useful information on the generated fragments, necessary for the elucidation of the structure of the ion selected in Q1.
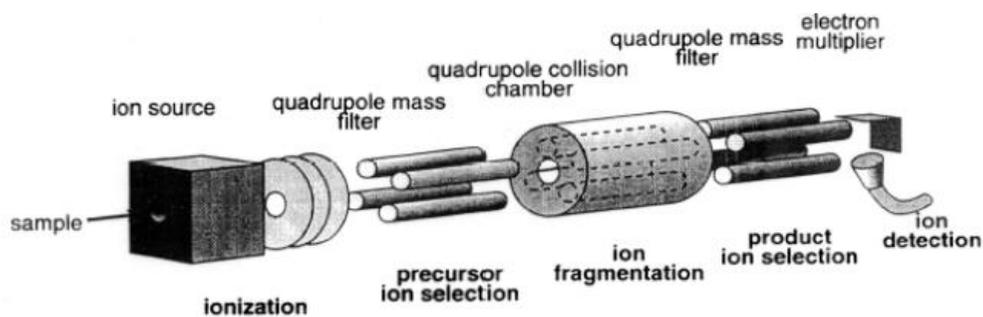
Figure 1.10: scheme of a triple quadrupole (QqQ) configuration.

To guide and focus the ions from the source to the detector, it is necessary to optimize the values of the potential applied within the quadrupole. In particular, in the 3200 QTrap^{TM} (Sciex instrument) the following electrical parameters must be optimized (Fig. 1.11):

- Declustering potential (DP): it is the potential applied at the end of the orifice plate. It is necessary for the ions fragmentation and the reduction of noise.

- Entrance potential (EP): it is the potential applied to the entrance in Q0. It is necessary for focusing the ion through Q0.

- Collision cell entrance potential (CEP): it is the difference of potential between Q0 and Q2. It is required for the flow of the ions trough Q2 (collision cell).

- Collision energy (CE): it is the potential applied between Q0 and Q2. It is fundamental for the fragmentation efficiency in Q2.

- Collision cell exit potential (CXP): it is the potential applied between Q2 and Q3. It aims to facilitate the ions transmission towards Q3.
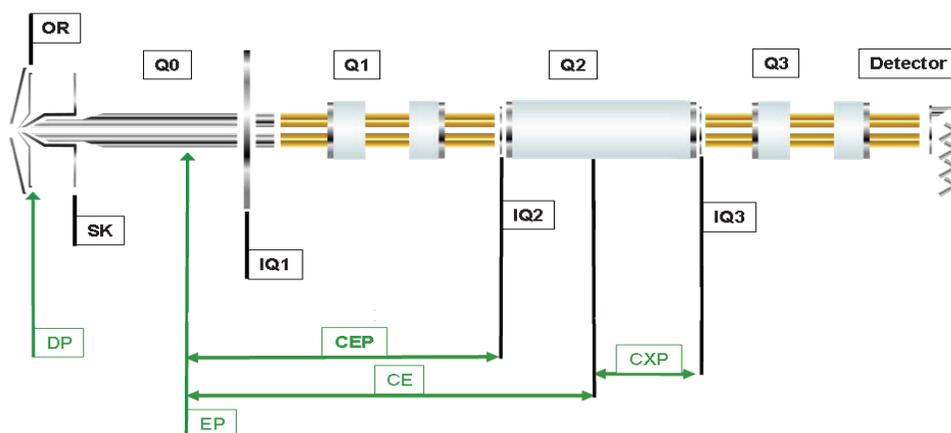
Figure 1.11: scheme of triple quadrupole electrical parameters.

Qualitative analysis are achieved using QqQ as a mass scan, provided by ramped DC and RF voltages, while quantitative analysis are performed using QqQ as a selective filter.

QqQ can operate in four different modalities: product ion scan, multiple reaction monitoring, neutral loss and precursor ion scan.

In *product ion scan* (MS/MS), Q1 selects the precursor ion among those generated in the ion source: the ions characterized by other *m/z* ratio are filtered out. Through

a process defined as Collision Activated Dissociation (CAD), Q2 generates product ions characteristic of the selected precursor ion by colliding with an inert gas, like nitrogen. The ions are accelerated by the electrical potentials, thus they increase their kinetic energy, being able to collide with the inert gas. During the collisions, part of the kinetic energy is converted into internal energy, leading to bonds breaking and molecular ion fragmentation. Q3 operates in scan mode, separating the ions characterized by different *m/z*. The overall result is a MS/MS spectrum of the *m/*z selected in Q1 (Fig. 1.12).

Product Ion Scan

MS1 static	Collision Cell RF only (pass all masses)	MS2 scanning
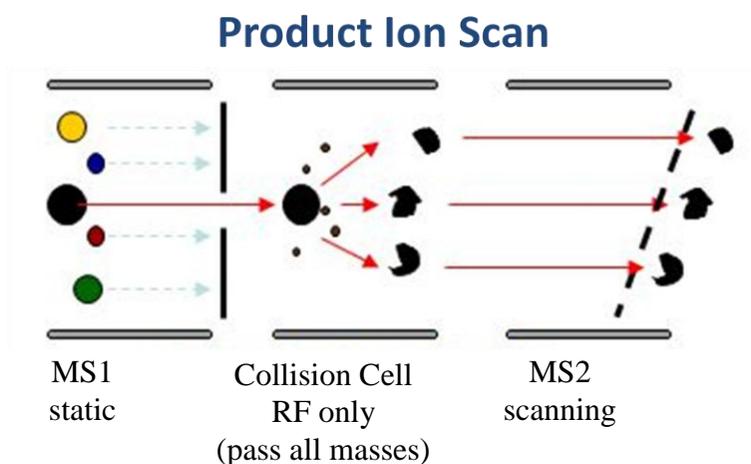
Figure 1.12: Product Ion Spectrum.

In *Multiple Reaction Monitoring* (MRM) Q1 selects a specific precursor ion and sends it to Q2 where it is fragmented (Fig. 1.13). Q3 acts again as a mass filter, transmitting only specific *m/z*, while the others are filtered out. The compounds typically show many transitions: the two most intense are respectively defined as *quantifier* and *qualifier*, used for the identification and confirm of a target analyte. The use of MRM mode allows the reduction of noise, providing larger signal/noise ratios (S/N), hence lower limits of quantitation (LOQ) and the increase of selectivity, because of the double filtering steps (in Q1 and Q3); it represents the best modality to perform quantitative analysis.

MRM: Multiple Reaction Monitoring

| MS1 static | Collision Cell RF only (pass all masses) | MS2 static |
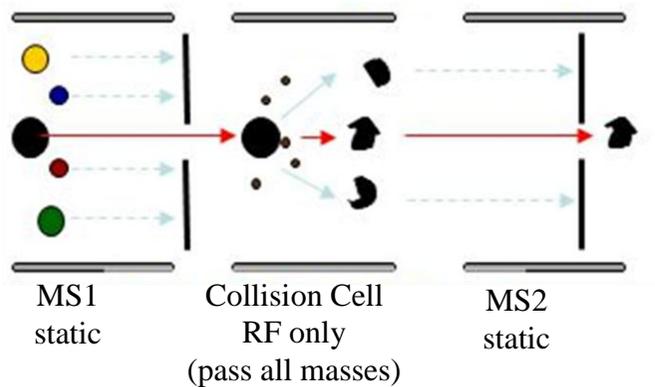
Figure 1.13: scheme of Multiple Reaction Monitoring.

In *neutral loss scan* (NL), both Q1 and Q3 work in scan mode, while the fragmentation occurs in Q2 (Fig. 1.14). The QqQ calculates the mass loss corresponding to Q1-Q3, finding peaks of the specified mass loss [63].

Neutral Loss Scan

| MS1 scanning | Collision Cell RF only (pass all masses) | MS2 scanning |
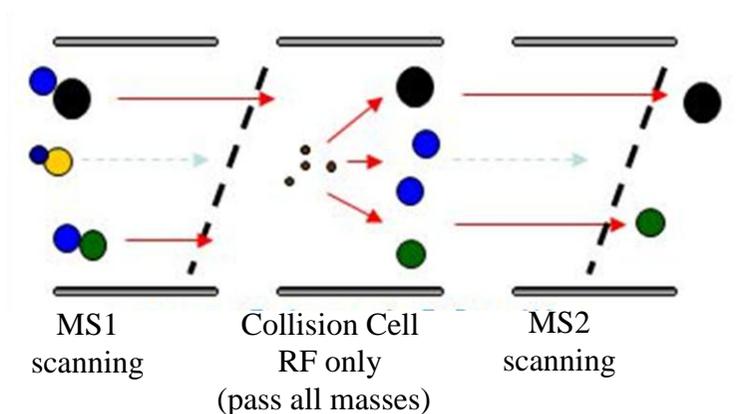
Figure 1.14: scheme of Neutral Loss.

In *Precursor ion scan* (PIS) mode, Q1 works in scan mode, Q2 acts as a collision cell, while Q3 is focused on a single product ion (Fig. 1.15). PIS experiment allows structurally similar compounds identification, since they generate a common ionic product.

Precursor Ion Scan

MS1 scanning

Collision Cell RF only (pass all masses)

MS2 static
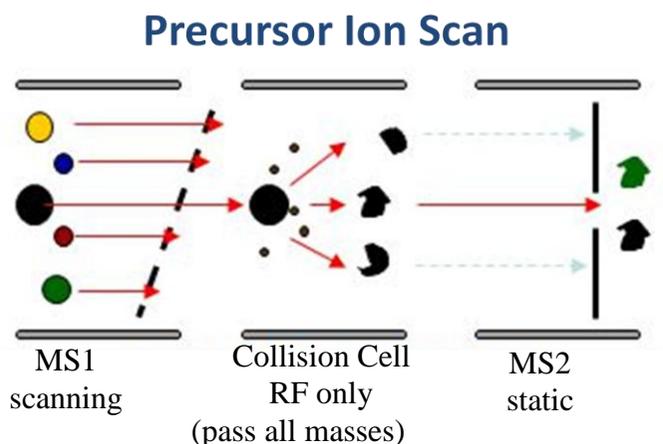
Figure 1.15: scheme of Precursor Ion Scan.

Thanks to the described features, QqQ is very suitable for the identification of many different target analytes in a single experiment. In particular, in this thesis, a hybrid Q-Trap analyzer working in MRM mode as QQQ analyzer was used for the determination of eight polyphenols and pantothenic acid in extra-virgin olive oil samples (Chapter 4) (Fig. 1.16)
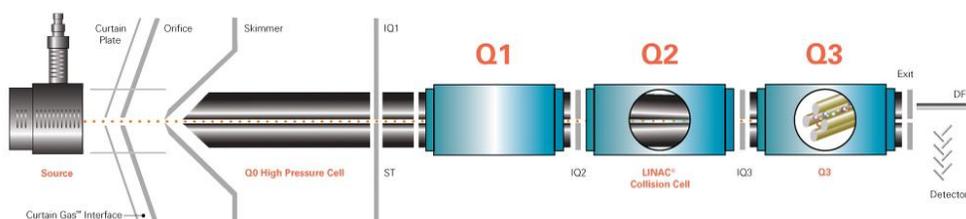
Figure 1.16: scheme of a QTrap (3200 QTrap, Sciex) configuration.

1.4.3.2 Time-of-flight

The general idea of the time of flight analyzer (TOF) dates back to the 40s, when Stephens [64] suggested that a mass spectrometer could be realized with an electrical field capable of separating group of ions on the basis of their different m/z ratio. Two years later, the first TOF instrument with a pulsed source, a free-field drift region and a revelation system was made by Cameron and Eggers (1948) [65]. However, this instrument did not generate any interest because of the low resolution provided and the imperfections of the electronics. Time-of-flight analyzers had to wait a long time for their development. In particular other innovations were necessary, like for example time lag focusing [66], delayed ion extraction [67-69], mass reflectron, allowed their main analytical performances to reach a resolution of over 20000 (full width half maximum, FWHM) and a mass accuracy below 50 ppm. These improvements allowed TOF to be considered one of the most powerful and versatile analyzer for analytical purposes [70-72].

The simplicity of the basic principles and the performances of the TOF analyzer can partly explain its current success. In TOF spectrometers working in "linear" mode, the ions generated by the source are accelerated through the application of an electric field (E) along a drift-free region of the field (flight tube L) (Fig. 1.17). Ideally, all the ions enter the flight tube with the same kinetic energy E_k:

$$E_k = zeE$$

(being z is the number of charges and e the electron electric charge) and

$$E_k = \frac{1}{2}\,mv^2$$

The speed of the ions (v) is independent from their masses (m),

$$v = \sqrt{\frac{2zeE}{m}}$$

Then the time (*t*) spent to cross the drift region is linked to the ion m/z value and can be recorded and converted into a mass spectrum:

$$L = vt$$

L being the distance between the ion source and the detector:

$$t = \sqrt{\frac{m}{2zeE}} \cdot L$$

hence

$$m/z = 2eE \cdot \left(\frac{t}{L}\right)^2$$

Typical times of flight ranges between 1 and 50 μs.

TOF mass-resolving power is hence related to time-resolving power by:

$$m/\Delta m = \frac{1}{2}\left(\frac{t}{\Delta t}\right)$$
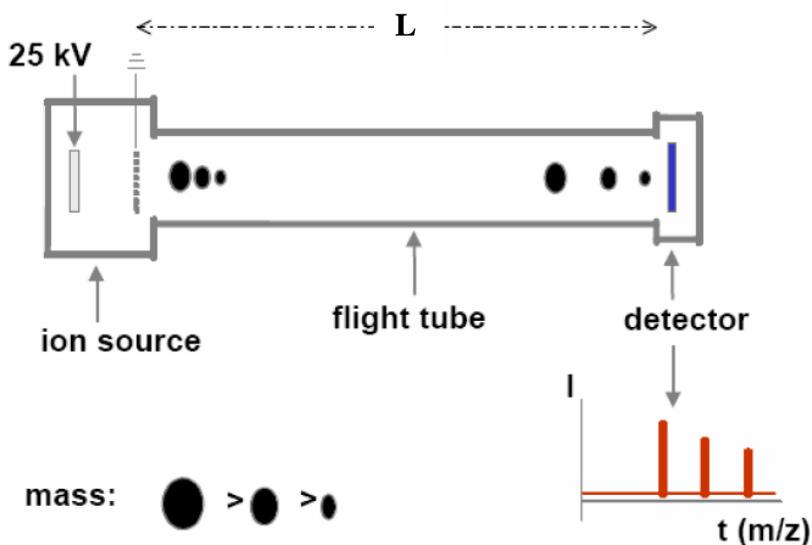
Figure 1.17: scheme of a mass spectrometer equipped with a time-of-flight (TOF) analyzer in linear mode.

Despite the simplicity of the principles underlying the TOF analyzers, the instruments equipped with flight tubes did not immediately reach success mainly for two reasons: they both showed low resolution and mass accuracy and were not easy to be interfaced with continuous ion sources. The initial conditions of ions formation (i.e. temporal and spatial conditions, and the distribution of kinetic energies), their fate during the fragmentation processes (metastable fragmentation) and the properties of the acquiring system (sampling rate and response of the detector) represented the most critical points. Ion spatial distribution was corrected by the use of two extraction regions that, changing length and intensity of the applied field, allowed the ion focusing close to the detector. A further improvement was given by the *mass reflectron* that minimizes energy dissipation. The reflectron (Fig. 1.18) is positioned at the end of the drift area before the field; it is constituted by a series of rings and/or grids that follow a linear gradient of potential, increasing (in positive modality) along the direction of motion. The ions cross the reflectron until they reach the part characterized by zero energy where they are bounced

through the reflectron itself, from which they exit with an energy equal to the one they had at the entrance. Isobaric ions (same integer mass, but different elemental composition, hence different exact mass) interact with reflectron on the basis of their speed, penetrating the electrostatic barrier proportionally to their kinetic energy increase. The more energetic ions penetrate more the reflectron, while the slower ions are deflected first. When the extra time spent by the faster ions in the reflectron is equal to the additional time spent by the slower ions in the primary acceleration region and in the drift tube, the maximum resolution is obtained. The mass spectra improvements that can be obtained thanks to the use of reflectron are reported in Figure 1.19. Thanks to such device, the resolution is increased up to 20.000-40.000.
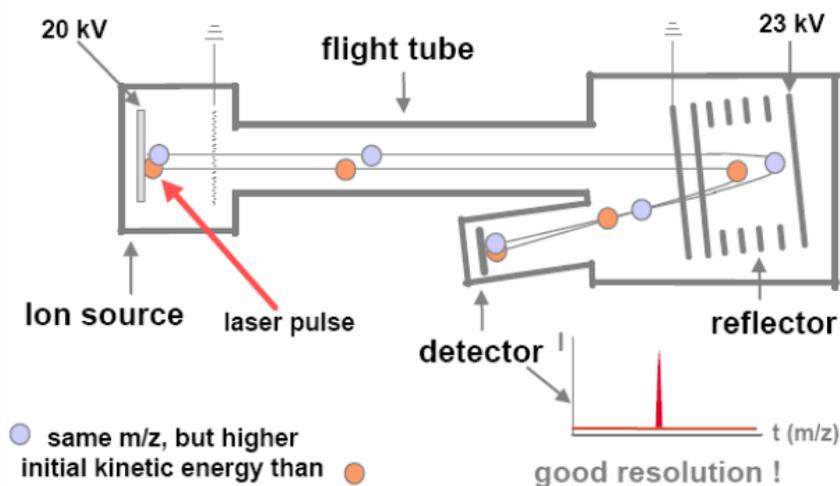
Figure 1.18: scheme of a mass spectrometer equipped with a time of flight (TOF) analyzer in reflectron mode.
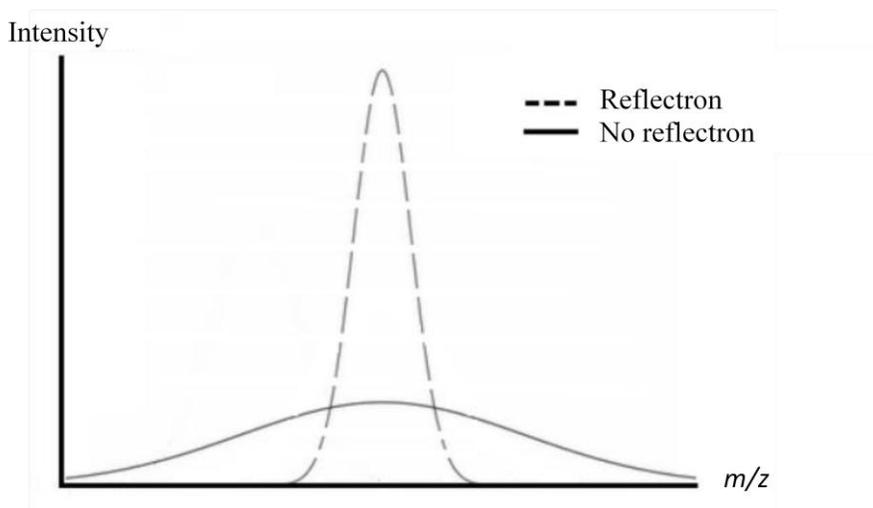
Figure 1.19: improvement of the analytical resolution obtained by the use of reflectron.

TOF instruments are characterized by robustness and, theoretically, they provide an unlimited mass range. They need software allowing fast data acquisition, since the groups of ions reach the detector at very short interval one from each other (in our instrument, the value is 25 ps). This permits to interface a continuum ionization source as ESI to the TOF analyzer.

1.4.3.3 Hybrid quadrupole-TOF (Q-TOF) analyzer

In recent years, there is a growing trend in substituting the triple-quadrupole mass spectrometers with Q-TOF instruments [73]. A Q-TOF consists of a quadrupolar mass filter (Q1), a collision cell (Q2) and a TOF. Figure 1.20 reports the scheme of the Q-TOF (5600 Triple TOF, Sciex) employed in this thesis for target and non-target analysis of all the investigated matrices.
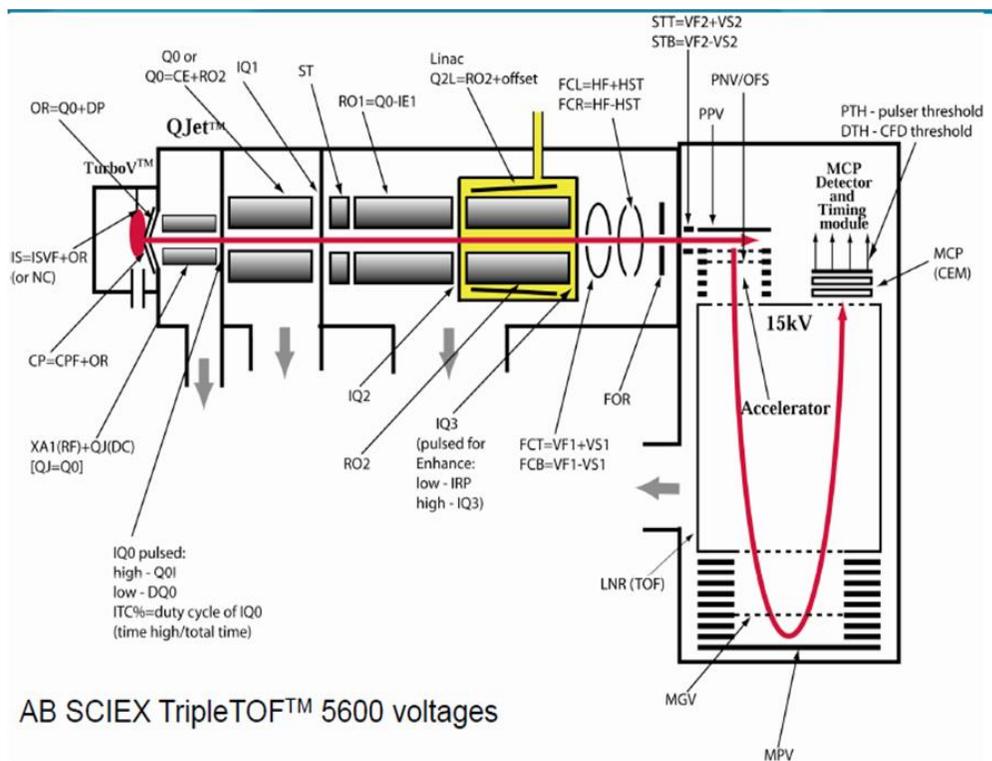
Figure 1.20: scheme of Q-TOF (5600 Triple TOF, Sciex).

The particular combination provides resolutions of 30.000-40.000, and an accuracy of few ppm. The main advantage of this instrument is due to the possibility of acquiring MS/MS spectra with high-resolution and accuracy. In this way, it is possible to obtain rapid and almost unequivocal molecular formulas (and structures) attributions to the molecular fragments. Figure 1.21 reports the scheme through which a group of ions originating from an analyte in Q1 is fragmented in the collision cell (Q2) and the corresponding product ions are detected in the TOF generating the corresponding MS/MS spectra.
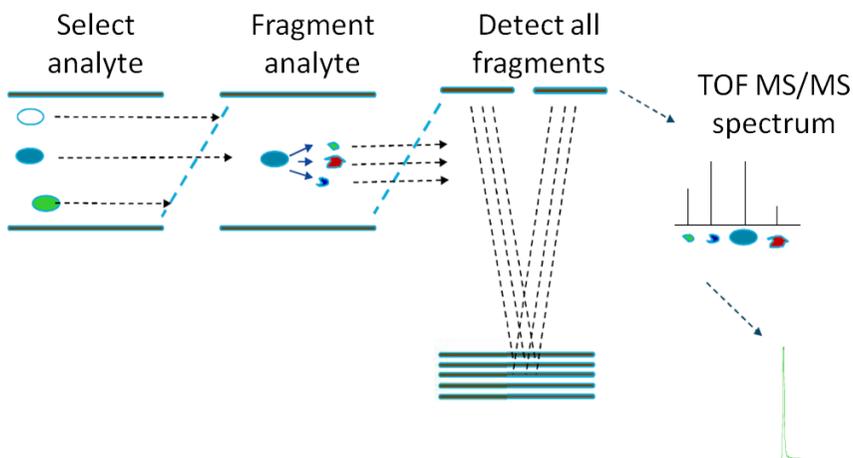
Figure 1.21: scheme of ions fragmentation in QTOF.

The use of a hybrid Q-TOF instrument with respect to the QqQ, allows faster scansion speeds, hence the possibility of performing sequential cyclic acquisitions. As it can be seen in Figure 1.22, the high acquisition speed allows the sampling of enough points of a chromatographic peak, providing quantification and structural identification through the acquired high-resolution and accuracy MS/MS spectra.
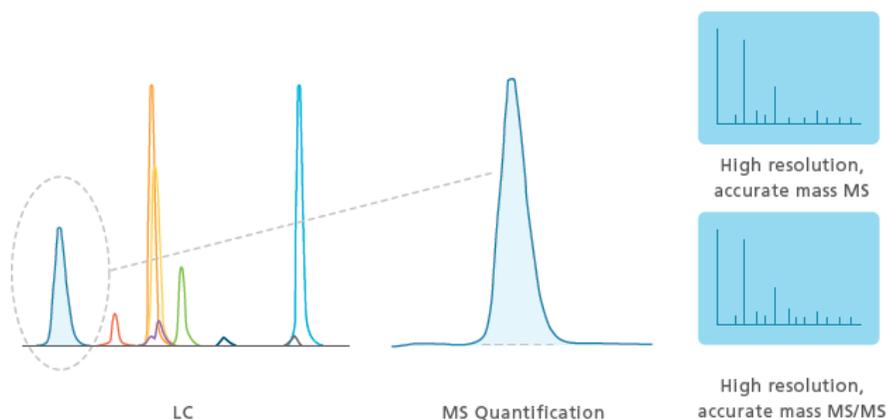
Figure 1.22: a high resolution MS/MS spectrum is acquired for each chromatographic peak.

QqQ instruments can be considered as the workhorses for target analysis since they provide high sensitivity and selectivity (especially when MRM mode is used). On the other hand, the unitary resolution and the low sensitivity in full scan mode make QqQ less suitable for non-target analyses than TOF, characterized by 20,000 as resolving power.

Hybrid instruments consisting in the combination of two different mass spectrometer types (such as QTOF) provide excellent detection and identification capabilities for low molecular weight compounds in complex matrices, based on high-resolution accurate mass measurement of precursor and product ions [74-78].

1.4.4 Detector

The most common detector is the electron multiplier (Fig. 1.23).
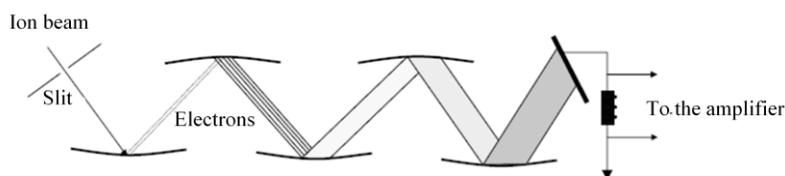
Figure 1.23: A discrete dynode electron multiplier.

This device is constituted by a series of dynodes kept at an increasing potential, in a way similar to that of photo-multiplier for detecting UV and visible radiations. The electrons are emitted in cascade from the cathodes and dynodes surfaces (in Cu/Be), because of the collisions with highly energetic ions or electrons. The ion beam is then accelerated ($\approx$5 KeV) after the collision on the cathode surface that emits electrons because of the photoelectric effect. These electrons are accelerated, attracted by the successive dynode (kept at a higher potential than the previous one), hence they collide with its surface where new electrons are so far generated, thus multiplying their number up to produce a measurable current.

41

For each ion hitting the cathode surface, a high number of electrons is produced on the last dynode. Furthermore, multipliers equipped with a maximum of 20 dynodes are commercially available and they are able to provide an increase of the current intensity of the order of 10^7.

Generally they are robust and reliable devices, characterized by a time of response of the order of nanoseconds.

Micro-Channel Plates (MCP) detectors are continuous-dynode photomultipliers in which the incoming signal is amplified throughout the whole length of the device. MCP is a planar component consisting of 104-107 parallel arrays of channels: each acting as a continuous-dynode electron multipliers. The diameter of each channel may vary according to the MCP model, but it ranges between 10 μm and 100 μm, while the thickness of each plate ranges from 40 μm and 100 μm (Fig. 1.24).
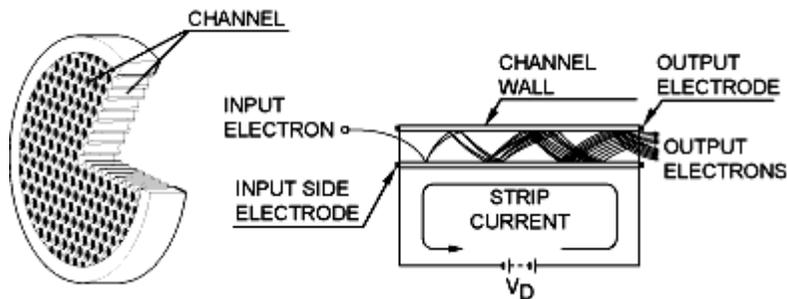
Figure 1.24: scheme of a single channel of an MCP.

The part that supports the channels of an MCP is usually made of leaded glass. The electric contact between the continuous dynodes is guaranteed by a metallic coating, usually made of Nichrome (a nickel, chrome and steel alloy) acting as the circuit cathode and anode in the front and back part of the MCP. The total electrical resistance between all the different electrodes is 10^9 Ω. Electrons multiplication depends, in particular, on the thickness and the diameter of the channels; typical

values range from 10^4 to 10^7 depending on the voltage applied between the two substrates. The main advantage with respect to the classic photomultiplier is represented by the typically short distance that the electrons have to cover before being collected on the anode. This makes it possible to obtain temporal resolutions generally lower than 100 ps, which characterize the rapid response of this detector. Potentially, spatial resolution is limited only by the channels size and their reciprocal distance (typically distant 3 μm for channels wide 12 μm). A positive voltage (V_D) is applied to each end of a single channel of an MCP, allowing the acceleration and multiplication of the initial electrons (Fig. 1.24).

A single incident particle (ion, electron, photon, etc.) enters a channel and emits an electron from the channel wall. A positive potential (of the order of hundreds of Volts) is applied between the photocathode and the ends of the MPC in order to efficiently collect the electrons produced by the photocathode. The produced secondary electrons hit the channel surface, multiplying their number. The process is reiterated throughout the channel length, leading to an electron cascade (thousands of electrons) emerging from the back of the plate.

1.4.5 Non-target data acquisition modes: Data Dependent Acquisition and Data Independent Acquisition

The MS/MS acquisition modes can be divided into two different approaches: the *Data Dependent Acquisition* (DDA) mode, the older and more employed, and the *Data Independent Acquisition* (IDA) mode, the most recent.

Through DDA mode, a first MS experiment (*survey scan*) is triggered with other MS/MS experiments (*dependent scans*), if prefixed criteria (*m/z* ion intensity, dynamic exclusion, background substraction) are met. The cyclical repetition of the survey scan and of the dependent scans during the whole duration of the LC-MS analysis, allows the acquisition of a great amount of information. As *m/z* abundance is the most important criteria that drives precursor ions selection in

DDA mode, such approach does not allow the complete exploration of all the ions present in the sample, as it is focused only on the most abundant ones, picking the ions in order of decreasing abundance (Fig. 1.25). The setting of the "dynamic exclusion" parameter among the DDA ones aims to "balance" the restrictive imposition of the m/z threshold value, since it avoids the reselection of possible precursor ions [62, 79].

Furthermore, DDA mode makes more improbable that a MS/MS spectrum of a precursor ion is acquired at the apex of its chromatographic elution profile, adversely affecting spectral quality [62, 80]. It also has to be underlined that DDA become less efficient in ion picking as the sample complexity increases, because the selection criteria worsen the limitations for identification and quantification [80-82].

To investigate all the ionizable compounds present in the sample, DIA mode is to be preferred to DDA. It consists in the fragmentation of all the detectable ions within a wide m/z range, without any exclusion criteria, providing a dynamic range of detected signals, better sensitivity and accuracy for the quantification of the investigated species, also increasing the reproducibility of the identification (Fig. 1.25) [80, 83].

The m/z ranges can be investigated adopting two different approaches. The first, *broadband DIA*, involves the fragmentation of all the ions that enter the mass spectrometer in a particular moment of the chromatographic time (All Ion Fragmentation, MS^E or MS^{ALL}), rolling low and high collision energy scan. The second one consists in dividing of the m/z ranges in smaller m/z isolation windows, independent one from each other and consecutively analyzed (Sequential Windowed Acquisition of All Theoretical Fragment Ion: SWATH, eXtendend Data-Indipendent Acquisition: XDIA, Multiplex data-independent acquisition: MSX, Precursor Acquisition Independent From Ion Count (PAcIFIC) [62, 84]. The selection of a unique isolation window, provided by the broadband DIA, allows to cover the whole precursor m/z range, to generate a large number of data points

during the elution that are necessary for an accurate quantification [83]. Despite DIA theoretically provides MS/MS information for all the possible precursor ions, it lacks in selectivity: this could represent a problematic aspect in the identification of the compounds [85]. Conversely, the subdivision of the precursor ion *m/z* range into smaller windows allows the improvement of the efficiency of fragmentation, reduces the complexity of the acquired composite fragment ion spectra as well as precursor mass tolerance (which can influence identification results) [83]. Furthermore, acquisition of high-resolution MS data helps to increase the identification confidence, as high mass accuracy precursor *m/z* values are detected and assigned post acquisition instead of using the center of the isolation window. In principle, large isolation windows are used to cover a wider precursor mass range with faster cycling rates or with extended acquisition times for each spectrum [83]. However, large isolation widths increase the number of precursors concurrently fragmented in the respective window, increasing the probability of overlap of fragment ions from different precursors (fragment ion interference), which also depends on the resolution and mass accuracy of the fragment ion signals [83]. For this purpose, narrower isolation windows are preferred to broadband DIA, since they increase the DIA dynamic range, as the most abundant species only affects a little part of the *m/z* range [62]. This approach was first developed in ion trap [86], Fourier transform ion cyclotron resonance (FT-ICR) [87], Orbitrap [88], and more recently in quadrupole-time-of-flight system, with the name of Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH) [89]. SWATH performs data-independent fragmentation of all precursor ions entering the mass spectrometer in 20-30 *m/z* isolation windows. The whole *m/z* range of interest is covered by a continuous stepping of the isolation window. This allows numerous repeated analyses of each window during the elution of a single chromatographic peak and results in a complete fragment ion map of the sample [62].

For a complete non-target characterization of complex matrices, DIA is more advantageous than DDA. In order to exploit all its strength, it is necessary to employ elaborate processing algorithms to investigate the composite or multiplex fragment ion spectra [83]. Convoluted or multiplexed MS/MS spectra are generated without explicit association between each single precursor and its corresponding fragments [83]. As a result, DIA needs more sophisticated data analysis post acquisition than DDA [83].
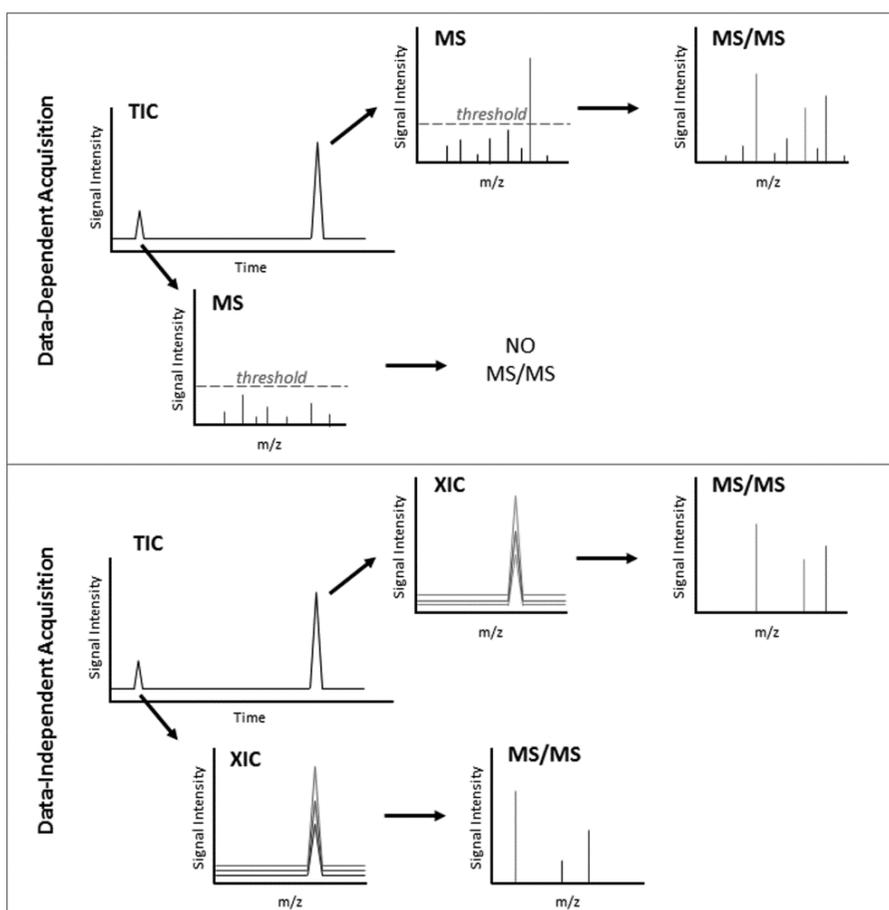
Figure 1.25: scheme of the difference between the data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes. In DDA, the selection of the precursor ion requires the passing of a certain threshold, in order to perform successive MS/MS acquisition. On the contrary, in DIA it is possible to acquire all MS/MS spectra of every precursor ions without any limitation.

1.4.6 Strategy for unknown identification

The identification of unknown compounds and their chemical structure elucidation is the most challenging aspect of non-target approach.

In order to propose reasonable putative structures it is mandatory to collect as much information as possible from the acquired data. Since the presented studies aimed to identify degradation products of compounds whose reference material was available, the starting point consisted in the parent compound mass spectrometry characterization. In fact, the acquired MS/MS spectra are useful for the identification of unknowns since some of the generated fragments and their patterns could be common to those belonging to the unknowns.

De novo identification of unknown compounds detected by MS/MS has the aim to correlate the observed spectral features in MS/MS to the structures of the investigated species, based on the reactivity of small organic ions during Collision Activated Dissociation (CAD) experiments. A good starting point for non-target identification is the analysis of standards available for the class to which the unknown can be supposed to belong, in order to collect information on their reactivity during CAD. For this purpose, the most common approach consists in the comparison between the fragmentation pattern of analogous species. Unfortunately, structural analogues different one from each other only for little moieties can be characterized by great differences of reactivity, that are reflected in their fragmentation patterns. A much more rigorous approach is to investigate how each structural modification influences reactivity, considering the chemical principles of resonance, inductive or steric effects, such as the availability of protons for abstraction. The interpretation of CAD reactivity starts from the site of protonation or deprotonation during ESI. It is important to underline that such site in solution does not necessarily correspond to the lowest energy position for ionization in the gas phase, as it is influenced by the media in which the reaction takes place. In fact, in solution the charges are stabilized by the solvent molecules, while in the gas

phase they are stabilized only by intramolecular effects (i.e. resonance, inductive effects, intramolecular hydrogen bonding). It follows that to assess the functional group with the greatest proton affinity (in positive ESI), it is not sufficient to consider the most stable proton site. Furthermore, ion dissociation during CAD occurs at vibrational excited energy levels, which often lead to charge transfer from the lowest energy structure. Currently, the best tool available for *de-novo* identification of unknown based on MS/MS spectral data is the understanding of the gas phase ion chemistry of small organic ions.

How can the unknown compounds be identified after an LC-MS/MS run? Generally, the first approach consists in a comparison between the chromatogram of a sample subjected to photoirradiation (degraded sample) with that of a sample not subjected to degradation (control sample) in order to identify photodegradation products [1, 4, 90, 91-98]. Figure 1.26 shows, as an example, the total ion chromatogram of a solution containing Sulforhodamine B dye before and after degradation.
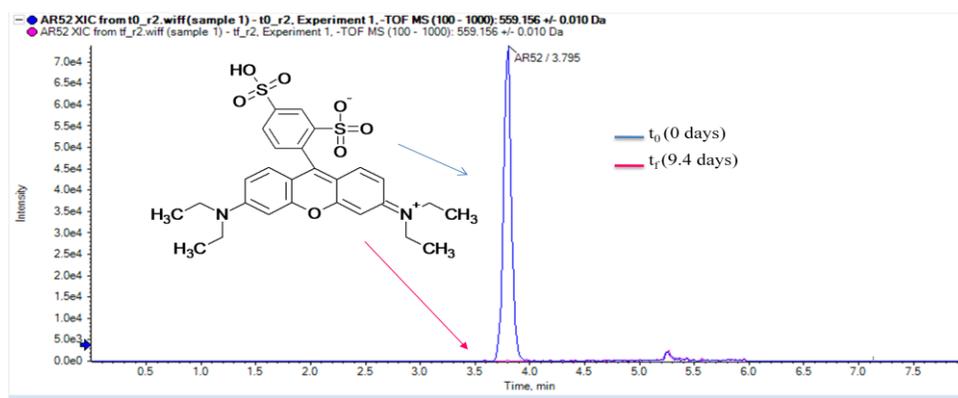
Figure 1.26: extracted Ion Chromatograms of the solution containing Sulforhodamine B dye before (blue line) and after (magenta line) degradation for 9.4 days under simulated sunlight irradiation. The chromatographic peak intensity of the dye decreases during degradation, up to totally disappear.

The setting of parameters such as approximate LC peak width (potential peaks that are less than 10% of this width are considered as noise, hence non selected),

background subtraction and minimum intensity threshold, allows the software to perform the semi-automated selection of non-target peaks along the chromatogram [80, 94]. When all the threshold parameters values are trespassed, the software compares the intensity ratio of the two chromatographic profiles, evaluating as significant only the peaks that trespass a prefixed threshold [80, 94]. Such workflow allows the selection of only the most significant peaks, hence their related *m/z* signals will be automatically included in a list by the software and successively extracted from the total ion current.

When using a low-resolution MS analyzer, the only way to propose chemical structures of unknowns is the concatenation of more MS experiments, such as full scan, enhanced resolution or zoom scan to evaluate the isotopic cluster with major accuracy, and MS/MS or MS^n, to obtain the widest information.

The correct matching between experimental spectral data for a specific compound and those present in a library database constitutes the simplest and fastest solution for unknowns identification. Spectral databases are commonly provided by instrument vendors with ESI-MS/MS instruments. Furthermore, it is possible to consult free online databases: Massbank, for example, contains more than 20,000 MS/MS spectra for thousands of different compounds [99]. Massbank is particularly appealing as it provides not only the MS/MS data from different instruments using different collision energies, but also merged spectra from a range of collision energies. In this way, it is possible to compare the experimental spectra with those present in the library, acquired at different collisional energies. Other examples of free online databases related to metabolites are the Human Metabolome Database [100] and Metlin [101]. NIST spectral database contains CID spectra of about 8,000 chemicals, collected using different ESI-MS techniques and instruments [102]. Unlike GC-MS, where the MS spectra are recorded at a fixed energy of 70 eV, in CAD no such universal choice of collision energy exists, because the collision energies required to induce dissociation depend on the ions mobility [80]. Furthermore, great differences occur between ion trapping

instruments capable of sequential multi-stage MS/MS (MS[n]) experiments and the radio frequency (RF) of quadrupole collision cells in hybrid instrument configurations like quadrupole/ion trap (QTRAP[TM]) and quadrupole time of flight (QTOF) instruments [80].

On the contrary, if the acquired spectrum of the unknown species is not found in a library, a meticulous interpretation of the spectrum is necessary.

In general, the chemical structures are proposed on the basis of:

- the molecular mass identified by the quasi-molecular ion present in the MS spectrum;
- the isotopic cluster;
- the assignment of even or odd number of N atoms corresponding to an even or odd molecular mass;
- the interpretation of the MS/MS or MS[n] analysis;
- the possibility of the species to ionize in both positive and negative modes;
- the plausibility of the retention time with respect to the calculated logP.

In most cases, these constraints strongly limit the number of possible structures [93, 95]. It is in this way that photodegradation products of sulforhodamine B [90] and those generated by the irradiation of three pesticides were identified using a hybrid analyzer (quadrupole time of flight). The possibility to use high-resolution MS analyzers permits to elucidate the chemical structures with much less uncertainty. Also in this case, in order to propose the chemical structure of the unknown species, it is necessary to gather the maximum amount of information; the most used are:

- the elemental compositions within the specific mass tolerance;
- the abundance of the isotopic cluster;

- the number of ring-double bonds (RDB);
- the use of accurate MS/MS spectrum, from which it is possible to obtain the elemental composition for each product ion;
- the plausibility of the retention time when correlated to the calculated logP of the unknown species.

High-resolution product ion MS/MS spectrum interpretation is of enormous importance in the elucidation of reliable precursor ions' molecular structures. The most relevant information used to elucidate the chemical structures of the unknowns is schematized in Figure 1.27.
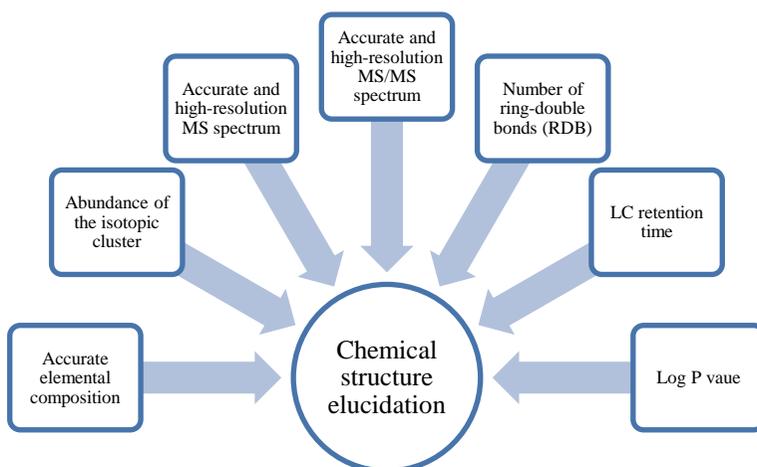
Figure 1.27: scheme of the main information useful for the chemical structure elucidation of unknown species.

Furthermore, some software allow the comparison of the MS/MS spectrum of the putative structure with the MS/MS spectrum derived by *in silico* fragmentation (i.e. simulated), providing as a result the percentage of matching or the number of common MS/MS signals [94, 95].

Mass Frontier (High-Chem, Thermo Scientific, San Jose, USA) and ACD Fragmenter (ACD/Labs, Toronto, Canada) are used for this purpose: they provide

highly specific identification of detected ions, on the basis of the interpretation of thousands of known reactions and fragmentation rules. An alternative approach is based on the generation of a list containing all the possible masses of the fragments formed by combinatorial break of the chemical bonds in a compound, evaluating the internal energy of each bond.

However, once a chemical structure is elucidated, it should be confirmed by the comparison with the reference material (if commercially available) or by complementary techniques like Nuclear Magnetic Resonance (NMR).

Howsoever, when the number of samples to compare is too high, the above described approach based on the direct comparison between the sample and the control chromatograms is unsuitable. For this purpose a *multivariate approach* should be preferred when a great amount of data is generated by the MS and MS/MS spectra of the considered samples. This approach is based on the use of multivariate statistical analysis, which is typically able to reduce the redundant information (Paragraph 1.5), extracting only the useful information, represented by the significant *m/z* signals related to the differences occurring between the analyzed samples. The multivariate statistical method exploited in this thesis are Principal Component Analysis (PCA), Principal Component Analysis-Discriminant Analysis (PCA-DA) and Partial Least Square-Discriminant Analysis (PLS-DA).

Chemometric approach will be described in detail in Paragraph 1.5.

1.5 Chemometric approach

Many of the problems related to the chemical field are defined almost daily, as *complex problems*. Such meaning is due to the extremely high number of variables and parameters necessary for the description of the considered experimental reality, to the possible synergistic or antagonists effects that may occur between the compounds present in the matrix that has to be analyzed, to correlation effects between variables and factors through non-linear effects. Furthermore, the experimental error makes mandatory the use of statistical tool in order to evaluate which factors are statistically significant, thus what is the relevance of each of the considered variables. It is clear how such kind of problems are functions of both the high number of variables and of the complexity of the analyzed problems, and they can only be managed through sophisticated statistical tools.

Chemometrics is a scientific discipline devoted to the development of statistical tools necessary to mine, elaborate and model the chemical data. Usually, multivariate problems are translated into numerical tables containing the scientific information (systematic variations), but also casual fluctuations, such as those related to the experimental error. The information can be useful or not, depending on what is the aim of the work. In other words, chemometric methods allow the separation of scientific information from the experimental error.

To do this, the data have to be organized in matrices: the X matrix reports the variables (indicated as *p*) that have to be studied on the lines, while the samples (indicated as *n*) are reported on the columns:

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,p} \end{bmatrix}$$

Experimental answers are also organized in a matrix (Y) where the parameters that have to be measured are indicated as *p'*:

$$\begin{bmatrix} y_{1,1} & \cdots & y_{1,p'} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \cdots & y_{n,p'} \end{bmatrix}$$

There are two different kind of approaches:

1. **One variable at a time (OVAT)**. It considers one variable at a time, but it does not highlight the possible correlations that can occur between the variables, providing only a partial perception of the problem;

2. **Multivariate**. It remediates the lack of the classical method, operating simultaneously on all the variables present.

Before performing a statistical analysis, it is necessary to follow a data pretreatment procedure in order to:

- complete any missing data and work on "complete" data matrices;
- to perform data transformations in the case of non-linear effects or in presence of particular trend or distribution and behavior of the data or of the experimental error distribution;
- to scale the data in order to remove scale effects occurring between the different variables.

Since most of the statistical techniques cannot operate on incomplete data matrices, it is necessary to manage the missing values, which can be done adopting different strategies:

- The elimination of the entire sample; it is a rarely used technique, especially when the number of available samples is low;
- The elimination of the variable; it is again rarely used since it cause the potential elimination of a great amount of information;
- The substitution of the value with the average of the variable; it is often adopted since it does not alter the structure of the data;

- The substitution of the missing value with a random one, chosen within the range valid for that variable;

- The substitution of the value through the regression method. It is adopted when it is possible to identify a variable highly correlated to the variable that present the missing value;

- The substitution of the value with the average of the class;

- The substitution through local similarity (*k-nearest neighbor, K-NN*); it is often adopted when the considered dataset contains a large number of variables that include missing values. Objects for which there are no missing values are used to estimate the distance from each object for which there is a missing data.

- The substitution through the method of principal components.

The next step is usually mandatory and consists in the scaling of the data in order to make the variables comparable to each other, thus avoiding any possible scale effect. Among the most diffused methods, there are: centering, autoscaling and range scaling.

In centering, the data are centered on the average value $\overline{x_J}$ of the variable through the transformation:

$$x'_{i,j} = x_{i,j} - \overline{x_J}$$

where the $x'_{i,j}$ represents the average value of the j-variable. In this way, the mean value of the new variable is zero and the variance is not altered.

Through autoscaling, the values are centered on the average value as in the previous case, but they are also normalized to unitary variance using the following equation:

$$x'_{i,j} = \frac{x_{i,j} - \overline{x_J}}{s_j}$$

The variables are comparable to each other for the scale factor because the variance s^2_j represents the amount of information brought by each variable: if all the variables have the same variance, they have the same amount of information.

Range scaling is a linear transformation that allows the data to lay between two desired values (usually -1,1 or 0,1).

1.5.1 Principal Component Analysis (PCA)

The term Principal Component Analysis refers to a multivariate technique proposed by Pearson in 1901 and developed successively by Hostelling in 1930. It allows to find out the existing correlations between variables, to efficiently visualize the objects, to resume the data description and reduce the dataset dimensionality. It is an unsupervised technique since no information related to the class structure of the analyzed data is used, except for their descriptors: it is the algorithm that looks for structured information sources in the data. The experimental variables are transformed into new variables, the so-called principal components (PCs), obtained as linear combinations of the original variables. The first component lies on the direction of the maximum variance, that is to say along the direction of the data maximum variability; the second one is aligned with the direction of the maximum residual variance and it is orthogonal to the previous one (Fig. 1.27):
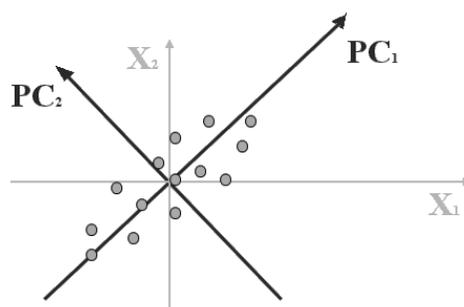
Figure 1.27: plot of the Principal Components.

56

The other following components contain a progressively decreasing amount of information: for this reason, the PCs are defined as hierarchical. Only the components that explain a relevant quantity of information useful for the solution of the problem are considered as significant.

From a mathematical point of view, the calculation of the main principal components can be performed in several ways, the most used concern the diagonalization of the variance-covariance matrix (i.e. singular value decomposition) or in the application of iterative algorithms (i.e. NIPALS). The variance-covariance matrix is a square matrix obtained from the following matricial equation:

$$\frac{X^T \cdot X}{n-1}$$

where X^T is the transposed X matrix and n represents the number of samples considered. The diagonal of the so obtained matrix is constituted by the variances of the variables, equal to 1 if the data were previously autoscaled; the extradiagonal values represent the variables covariances that range from -1 to 1 in the case of autoscaled data.

The diagonalization of the matrix variance-covariance leads to a new diagonal matrix characterized by extradiagonal values all equal to zero since the principal components are orthogonal, and diagonal values indicated as λ_1, λ_2, λ_3,... λ_p, are related to the quantity of information explained by each principal component, hence since they are hierarchical, also these values are hierarchical ($\lambda_1 > \lambda_2 > \dots \lambda_p$). The so obtained matrix is defined as the matrix of the eigenvalues:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}$$

Defining with the X (n,p) the matrix of the orthogonal data constituted by *n* objects and *p* variables, with L (p, M) the matrix of the loadings where *p* represents the variables and *M* represents the number of principal components, the matrix product of these two matrices originates a new matrix containing the projection of the objects along the *M* considered principal components: T (n,M) or scores' matrix, constituted by the *n* objects described by the *M* principal components:

$$T = XL$$

$$(n, M) = (n, p)(p, M)$$

Each of the element present in T are calculated as:

$$t_{jk} = \sum_{i=1}^{p} x_{ji} \cdot l_{ik}$$

where l_{ik} is the loading of the *i*-variable in the *k*-principal component, x_{ji} is the value of the *i*-variable assumed by the *j*-object, while t_{jk} is the score of the *j*-object on the *k*-principal component.

It can be concluded that:

- the variance of the *m*-component is expressed by the corresponding eigenvalue;
- two different principal components are orthogonal, hence independent one from each other, with null co-variance;
- the PCs' variance have decreasing values, with the last components that contain the experimental error;
- total variance is the eigenvalues' matrix trace.

As previously described, principal components are obtained as linear combinations of the original variables and the loadings represent the coefficient of the linear combination:

$$PC_1 = l_1^1 x_1 + l_2^1 x_2 + .. + l_p^1 x_p$$
$$PC_2 = l_1^2 x_1 + l_2^2 x_2 + .. + l_p^2 x_p$$
$$PC_M = l_1^M x_1 + l_2^M x_2 + .. + l_p^M x_p$$

By definition, the following relationship takes place:

$$\sum_i (l_i)^2 = 1$$

Because of this equation, the loadings show values in the range between -1 and 1, and identify the direction of the principal component in the space of the original variables. The positive or negative value is not important for the evaluation of the information explained by the variable: only the absolute value has to be considered; the plus or minus sign allows the evaluation of which variables are positively or negatively correlated among them. The plot reported in Figure 1.28 is called the *loading plot* and it is the graphical representation of the loadings; in order to interpret it correctly, it is important to evaluate the position of the variables with respect to the origin of the Cartesian axes.
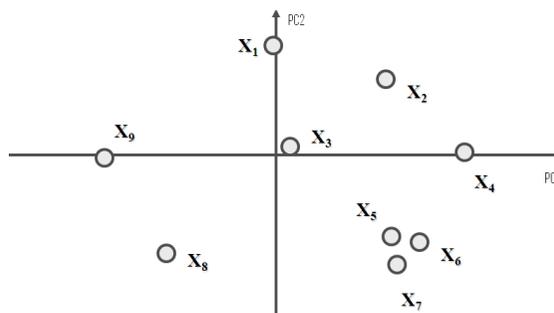
Figure 1.28: loading plot of the first two Principal Components.

For example, note that X_4 and X_9 are in the opposite position respect to the origin of the axes along PC1. This means that they are negatively correlated: if X_4 decreases, X_9 increases (this is true because of the percentage of the variance explained by the first PC). X_2 and X_8 are in the opposite position respect to the origin of the axes, but they both show a contribution on PC1 and on PC2. The variables X_5, X_6 and X_7 are very close one to each other, hence they are similar variables, directly correlated to each other; such condition implies that the three variables are redundant since they provide almost the same information. X_3 is positioned at the origin of the axes, i.e. it is not explained by the first two principal components, thus the information related to this variable will be considered by other successive PCs.

To exhaustively analyze the problem, it is also necessary to evaluate the *score* plot, or projection graph, that represents the object projections into the new space of the principal components [103].

The scores do not have values in the range between -1 and 1, therefore the scores analysis should be performed in parallel with the loadings:
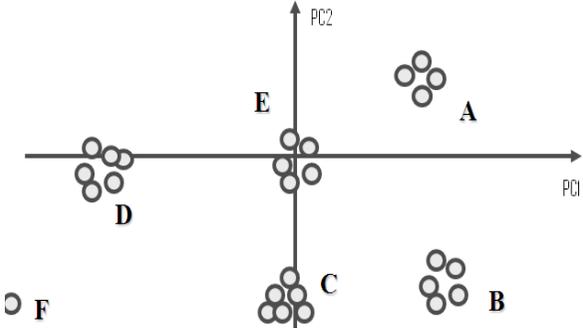
Figure 1.29: score plot.

As reported in Figure 1.29, the score plot contains the objects clustered in different zones and the scores that are near to each other highlight objects with similar characteristics.

Analyzing both graphs at the same time, it can be stated that variables X_5, X_6 and X_7 show high values for group B objects (if the variables were, for example, concentrations of analytes, those experiments would represent high concentration values for those particular analytes); group A objects show high values for X_2 variable and low for X_8.

There are few methods to choose which are the significant components, hence which have to be chosen to exhaustively represent the problem. The most adopted methods are that of the screen plot and the average eigenvalue one. The former is a graphical method that reports the percentage of the explained variance for a component as a function of the number of the principal components; the explained variance percentage is calculated as:

$$\% \, variance = \frac{\lambda_i}{\sum_{i=1}^{nPC} \lambda_i} \cdot 100$$

Normally, the cumulative variance is considered and it is calculated as:

$$\% \, variance \, cum. = \frac{\sum_{m=1}^{M} \lambda_m}{\sum_{m=1}^{p} \lambda_m} \cdot 100$$

The so obtained plot is reported in figure 1.30:
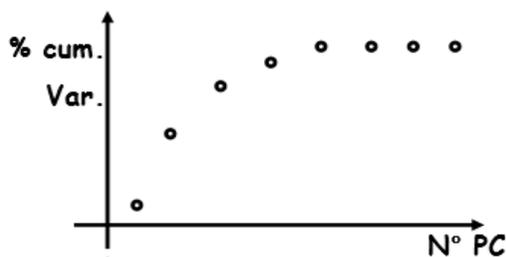
Figure 1.30: the plot reports on the Y axes the percentage of the cumulative variance explained by the principal components.

The significant component are those associated to the discontinuity point. Alternatively, it is possible to evaluate a parameter R defined as:

61

$$R = \frac{\lambda_m}{\lambda_{m+1}}$$

as a substitute of the cumulative variance. In this case (Fig. 1.31), the obtained plot has an opposite trend respect to the previous one:
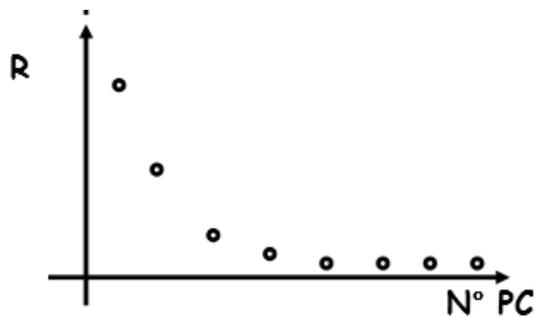
Figure 1.31: the plot reports the R parameter of each of the principal component.

The average eigenvalue method considers as significant the PCs that have $\lambda_i > \bar{\lambda}$ values if the data are not previously autoscaled. If they are autoscaled it is that $\bar{\lambda} = 1,$ thus are considered as significant all the principal components that have $\lambda_i > 1$. Each of the principal component explains, at least, the same of information of one independent original variable.

1.5.2 Classification methods

Classification methods are defined as *supervised* since the information related to the presence of the classes in which the objects are grouped, is defined at the beginning. Through classification methods, it is possible to attribute the objects to the existing classes, on the basis of the values of a group of variables or descriptors X. The classification methods can be divided into modeling and non-modeling ones. The formers offer classification methods that allow the definition of the area described by the class and the evaluation of the descriptors effects on the belonging

class. Conversely, the latters do not provide a model: they just group the objects in the available classes.

A classification method is a mathematic equation able to correlate the different variables (descriptors) with a Y variable that provides the belonging of each object to a class *a priori* defined. The two different models that can be proposed are deterministic or stochastic. The former is based on the presence of *a priori* hypotheses about the existing relationships between chemical, physical or biological properties of the system under analysis. The latter is not based on *a priori* hypotheses, it exploits only the information related to the general shape of the model and derive its conclusions from pure statistical analyses.

In this work of thesis, only stochastic methods are exploited.

The workflow that allows the realization of a model is articulated in different steps:

- decision of what type of model has to be adopted (deterministic, stochastic, linear, non linear...);
- realization of the model: numerical values are attributed to the model coefficients through the fitting procedure;
- model validation: the model is tested and applied in controlled conditions; the model has to be modified in order to optimize its fitting attitude (it has to be able to explain the studied system) and its predictive abilities when applied to unknown scenarios.
- application of the model to unknown scenarios.

The method validation is necessary since it is fundamental that they are able not only to explain the experimental scenarios used for their development (fitting), but also to generalize them to new scenarios (prediction) avoiding at the same time to model the experimental error (overfitting). For this purpose, the data are divided into two parts: one is used for the construction of the model (*training set*), while the remaining data (*evaluation set*) are used for the evaluation of the model predictive capabilities. Validation techniques differ according to the partition of the

objects into training and evaluation sets. The most adopted validation techniques are: the leave-one-out method (LOO) and the leave-more-out one (LMO). The former implies the exclusion of a single object at a time during the model building and its prediction with the model built in its absence; the model coefficients are hence calculated without the missing value. This will be used to evaluate the predictive abilities of the model itself [104].

LMO method is based on a simultaneous exclusion of a group of data from the model.

The two described models are different for the perturbation amount they cause on the calculated method: LOO is the less perturbative one since it eliminates one object at a time and it provides prediction performances still very similar to the fitting ones; LMO is the most perturbative one and it can be applied only when there is a consistent number of objects. This results in a different severity of the validation procedure.

The classes are usually defined on the basis of known characteristics, related to the problem that has to be faced. The classes can be qualitatively determined through the discretization of a continuum variable or using as classes the groups evidenced by principal component analysis, or by cluster analysis. Classification methods can be divided into three different categories:

1. Modeling methods. Examples of modeling methods are the Bayesian ones based on the Bayesian rule, i.e. LDA (Linear Discriminant Analysis), QDA (Quadratic Discriminant Analysis), NMC (Nearest Mean Classifier);

2. Non-modeling methods where the attribution of an object to a class is evaluated through the calculation of its distance from that class. The most popular method is KNN (K-Nearest Neighbours);

3. Tree diagrams, such as CART and LDCT methods.

Classification performances can be described through the classification or confusion matrix: the lines report the real classes, while the columns report the

assigned one. The diagonal elements represent the correct attribution, while the extra-diagonal ones represent the wrong attributions. The goodness of the classification can be described using the Non-Error-Rate% (NER%):

$$NER\% = \frac{\sum_g c_{gg}}{n} \cdot 100$$

It expresses the ratio of the correct attribution, in fact it is calculated as the ratio between the sum of the diagonal elements of the confusion matrix and the total number of the objects. The NER% can in turn be calculated in fitting or prediction.

1.5.2.1 Linear Discriminant Analysis (LDA)

LDA is a Bayesian method, based on the hypothesis that all the variance-covariance matrices of the classes are equal to each other and coincide with the pooled variance matrix, that is represented by the average variance-covariance matrix weighted for the degrees of freedom. An average shape of the classes is evaluated; the separation of the classes is given by a hyperplane and the classification or discriminant direction is a line perpendicular to the hyperplane. Along the discriminating direction, the distance between an object and its class is small, while it is large between the same object and the other classes.

An object x is assigned to the class g for which the posterior probability P(g/x) is maximum, assuming a Gaussian multivariate probability distribution:

$$P(g|x) = \frac{P_g}{(2\pi)^{p/2} \left|S_g\right|^{1/2}} e^{-0.5 \, (x-c_g) S_g^{-1}(x-c_g)}$$

where P_g is the prior probability of class g, S_g is the covariance matrix of class g that, in the case of LDA, is approximated with the pooled (between the classes) covariance matrix, c is the centroid of class g, p is the number of descriptors. The

argument of the exponential function is the Mahalanobis distance between the object x and the centroid of the class g, which takes into consideration the covariance structure of the class:

$$(x - c_g)S_g^{-1}(x - c_g)$$

From the logarithm of the posterior probability, by eliminating the constant terms, each object is classified in the class g for which it is minimum, the so-called discriminant score:

$$D(g|x) = (x - c_g)S_g^{-1}(x - c_g) + ln|S_g| - 2lnP_g$$

The selection of the variables of the LDA models which discriminate the classes present in the dataset was performed by a stepwise algorithm in forward search ($F_{to-enter}$=4.0).

1.5.2.2 Partial Least Square-Discriminant Analysis (PLS-DA)

PLS is a multivariate method, allowing the establishment of a regression model between a response variable (Y variable) and a set of descriptors (X variables).

Partial Least Square models both the X- and the Y- variables simultaneously, to find the latent variables in X that will predict the latent variables in Y. These PLS-components (latent variables) are similar to the principal components. If there are several responses, they are modeled together, in a multivariate way [105-107].

The predictive ability obtained from the leave-one-out or leave-more-out cross-validation on the training set is often used to find the optimal number of model components for PLS.

PLS could be also used as a classification method substituting the continous variable response Y with a Y variable that defines the belonging to the class: -1 is

attributed to the samples of the first class, while +1 is attributed to those of the second class. This method is called PLS-Discriminant Analysis (PLS-DA).

The coefficient of multiple determination R^2 for PLS models is calculated as:

$$R^2 = 1 - \sqrt{\frac{\sum_{i=1,n}(\acute{y}_i - y_i)^2}{\sum_{i=1,n}(y_i - \bar{y})^2}}$$

where the two sums run on the samples of training (R^2), test or production set (R^2_{CV}), $\acute{y}_i$ is the predicted value of the response for the *i-th* experiment, $\bar{y}$ is the average response of the training set for the calculation of R^2, while it is the average response of the test set when the production or test set is considered for the calculation of R^2_{CV}.

1.5.3 QSAR descriptors

The language we adopt to express concepts and to describe reality does not coincide with reality itself, but it plays an important role in shaping the reality we intend to speak about. The interpretation of each "fact" is therefore indissolubly linked to the ways in which the fact is described through the mediation of the language and meaning of every used term, which depends on the theoretical context in which it is adopted.

In this contest the descriptors are the words of the language through which we represent the studied chemical, physical or biological object. The type of descriptors used in modeling relationships between activity/property and molecular structure is thus decisive in determining the quality of the obtained patterns. Such relationships are known as QSAR (Quantitative Structure Activity Relationship) [108].

The importance of the problem is extremely relevant if we consider the hypothetical use of a new compound: it would be of great importance to know

some properties (chemical-physical or pharmacological) that can be predicted before the synthesis of the compound itself or before the beginning of an extensive experimentation. For example, the availability of a reliable estimate of the solubility, melting point, boiling point, surface tension, a defined biological activity of the ability to penetrate lipid membranes of a compound, using known properties of the compound itself, would be of great importance to the formulator. The most common descriptors classification, according to Hansch's approach, refers to three fundamental properties groups:

1. Hydrophobicity (such as Log P);
2. Electronica parameters (such as HOMO and LUMO energies);
3. Steric parameters of shape and size (such as molecular weight).

The molecular descriptors can:

- be generated by experimental measurements, theoretical calculations or by simple counting or summation operations;
- represent the entire molecule or a part of it or one of its substituent defined in a particular position;
- require the 3D molecular structure or its molecular graph or only its molecular formula;
- be defined by a scalar, a vector, a scalar field, and so on.

Molecular descriptors can be divided, for example, in one-dimensional (1D), bi-dimensional (2D), three-dimensional (3D) descriptors, microscopic or macroscopic, etc.

The 1-D descriptors are the simplest one, directly derived from the molecular formula of the compound; this is the case of scalar vectors such as molecular weight, the number of Carbon atoms or the number of bonds. The 2-D level descriptors derive from the molecular graph and the connectivity matrix, such as the topological descriptors. The 3-D level descriptors come from the three-

dimensional structure of the molecule, that is based on the knowledge of its actual spatial coordinates. Referring to the microscopic world, all descriptors that describe the conformational aspects of molecules are considered while, in the macroscopic world, the considered descriptors are the geometric ones, such as molecules volume and surface. Increasing the investigation depth to the microscopic level, means the considering of the electronic level, where descriptors are the energies of molecular orbitals, electronic density maps, quanto-mechanical parameters, and so on. At last, there are descriptors that also take into account interactions with the surrounding area, as most of the experimentally measured quantities do.

Molecular descriptors constitute an important group of different descriptors (chemical, physical, biological, environmental, geographical and geological, etc.) used to create mathematical models.

QSAR relations are mathematical models used in my case for the prediction of toxicity, based on the compounds physical characteristics related to their structures (molecular descriptors). Acute toxicities (such as the concentration that causes 50% mortality of individuals) are an example of the toxicities evaluation that can be predicted by QSAR models.

QSAR models calculate compounds toxicity using a linear function of the molecular descriptors, such as:

$$toxicity = ax_1 + bx_2 + c$$

where x_1 and x_2 are orthogonal descriptors, while a, b and c are fitting parameters. Molecular weight and partition coefficient octanol/water are examples of molecular descriptors.

Toxicities evaluation obtained from QSAR models can be adopted to evaluate non-tested compounds in order to propose the priority of the biological tests (expansive and time-consuming) designed to establish toxicities levels. If traditional biological tests were unavailable, QSAR models are a good alternative. Furthermore, QSAR

models are useful for the evaluation of toxicity, necessary for the design of *green processes* algorithms, such as the *Waste Reduction Algorithm* [109].

Using the software, it was possible to obtain a mutagenicity estimation of the degradation products, simulating an Ames test, according to which a compound is positive to the test if it induces a reversion in the growth of a *Salmonella typhimurium* colony strain.

In this work of thesis, QSAR models were adopted to evaluate the toxicity of the identified degradation products. The adopted procedure was:

- *Hierarchical method*: compound toxicity is estimated using the weighted average of predictions coming from different models. Such models are obtained using the Ward method to divide the training set into groups of similar clusters; a genetic algorithm is then used to generate models for each cluster.

- *FDA method*: the prediction for each compound is performed using a new model that has to be similar to the compounds structurally more similar to those that have to be evaluated. The predictions for each compound are then performed using a single model that contains compounds structurally similar to that intended to be analyzed, selected from the training set. It is opposed to the hierarchical method in which predictions are made using one or more clusters *a priori* built using the Ward method.

- *Nearest neighbor method*: toxicity is estimated considering the average of the three compounds in the training set structurally more similar to the compound to be analyzed. In order to make a prediction, any structure considered analogous must exceed a set similarity value, SC_{min}. The threshold value is set to 0.5 so that the prediction coverage is similar to that obtained from the other QSAR methodologies. Through Nearest neighbor method it is possible to obtain rapid external toxicity

evaluation (the compound that has to be tested is never present in the set of considered analogous structures). The drawback is that structural differences occurring between the compound that has to be tested and the analogous structures are not all considered.

- *Consensus method*: toxicity is estimated through the average toxicity predicted through the three QSAR methods previously described. If only one QSAR method is able to make predictions, the predicted value is considered unreliable and unused. This method offers the utmost precision of prediction as incorrect predictions are overcome by predictions from other methods. In addition, this method allows maximum prediction coverage as some methods with limited applicability domains are also evaluated.

Table 1.1 reports the main strengths and drawbacks of each of the proposed methodologies.

Method	Strengths	Drawbacks
Hierarchical	It provides reproducible predictions since it is the result of multiple methods.	Does not provide external evaluation of toxicity for the compounds present in the training set.
FDA	It can generate new models based on the nearest analogous to the compound that has to be evaluated and always provides an external toxicity evaluation.	Sometimes, predictions take time since a new model has to be calculated at each cycle.
Nearest neighbor	It allows a rapid toxicity esteem.	It does not use a QSAR model to correlate the differences between the compound of the test set and the similar compound.
Consensus	It provides the best predictive results during the external validation test.	It does not allow external toxicity evaluation for compounds present in the training set.

Table 11: strengths and drawbacks of the models adopted for toxicity evaluation.

Once the model iteration is completed, its goodness ($q^2_{adj,LOO}$) is calculated as follows:

$$q^2_{adj,LOO} = 1 - \left[\frac{\sum_{i=1}^{n_k} (\hat{y_i} - y_{exp,i})^2 / (n_k - p - 1)}{\sum_{i=1}^{n_k} (y_{exp,i} - \overline{y_{exp}})^2 / (n_k - 1)} \right]$$

The model is considered valid if the result of the equation is greater or equal to 0.5. If it is not the case, it will not be considered for the further prediction of the toxicity of the compound that has to be tested. In the Ames test, the model validity is determined by the LOO concordance, instead of the LOO q^2. Concordance is the fraction of compounds that are correctly attributed (for example, experimental active compounds that are predicted to be active and experimental inactive compounds predicted to be non-active). The model is considered valid if the LOO concordance is greater or equal to 0.8. Furthermore, both LOO sensitivity and specificity have to be at least 0.5 to avoid the use of models that can only partly predict the activity or inactivity of a compound. Sensitivity is the fraction of compounds that are experimental active and that are predicted to be active. Specificity is the fraction of compounds that are experimental non-active, and predicted to be non-active. To assess if a compound is toxic or not (active/non-active) the accuracy of prediction has to be evaluated in terms of fractions of compounds that are correctly predicted, thus accuracy prediction is defined as a function of concordance, sensitivity and specificity.

References

1. Gosetti F, Chiuminatto U, Mazzucco E, Mastroianni R, Marengo E (2015) Food Chem 167: 454-462.

2. Gosetti F, Chiuminatto U, Mazzucco E, Calabrese G, Gennaro M C, Marengo E (2012) Anal Chim Acta 746: 84-89.

3. Gosetti F, Mazzucco E, Gennaro M C, Marengo E (2016) Environ Chem Lett 14: 51-65.

4. Gosetti F, Chiuminatto U, Mazzucco E, Mastroianni R, Bolfi B, Marengo E (2015) Environ Sci Pollut Res 22: 8288-8295.

5. Bottaro M, Frascarolo P, Gosetti F, Mazzucco E, Gianotti V, Polati S, Pollici E, Piacentini L, Pavese G, Gennaro M C (2008) J Am Soc Mass Spectrom 19: 1221-1229.

6. Knolhoff A M, Croley T R (2016) J Chromatogr A 1428: 86-96.

7. Díaz R, Ibáñez M, Sancho J V, Hernández F (2012) Anal Methods 4: 196-209.

8. Motoyama A, Kihara K (2017) Mass Spectrom 6: 1-9.

9. Zhong Z, Li G (2017) J Sep Sci 40: 152-169.

10. Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products. O. J. EU, L342, 2009, 59/209.

11. Banned ingredients. http://www.cirs-reach.com/news-and-articles/the-technical-safety-standards-for-cosmetics-(2015)-will-come-into-effect-by-dec-2016.html .

12. Pellegrini M, Marchei E, Pacifici R, Rotolo M C, Pichini S (2011) J Pharm Biomed Anal 55: 842-847.

13. Bocca B, Pino A, Alimonti A, Forte G (2014) Regul Toxicol Pharmacol 69: 447-467.

14. Zakaria A, Ho Y B (2015) Regul Toxicol Pharmacol 73: 191-195.

15. Volpe M G, Nazzaro M, Coppola R, Rapuano F, Aquino R P (2012) Microchem J 101: 65-69.

16. Borowska S, Brzóska M M (2015) J Appl Toxicol 35: 551-572.

17. Thyssen J P, Menne T (2010) Chem Res Toxicol 23: 309-318.

18. Centers for Disease Control and Prevention, Mercury exposure among household users and nonusers of skin-lightenning creams produced in Mexico, California and Virginia (2012) Morb Mort Wkly Rep 61: 33-36.

19. Jackson L S (2009) Agric Food Chem 57: 8161-8170.

20. Moore J C, Spink J, Lipp M (2012) J Food Sci 77: R118-R126.

21. Malik A K, Blasco C, Picó Y (2010) J Chromatogr A 1217: 4018-4040.

22. FAO/WHO Food Standards, FAO and WHO, Rome, Italy, 2009.

23. Commission of the European Communities, White paper on Food Safety, COM (1999) 719 final, Brussels, 12 January 2000. Available at: http://ec.europa.eu/food/food/intro/white_paper_en.htm

24. Herberer T (2002) Toxicol Lett 131: 5-17.

25. Terns TA, Kreckel P, Mueller J (1999) Sci Total Environ 225: 91-99.

26. Rosen R (2007) Curr Opin Biotechnol 18: 246-251.

27. Li C M, Chu R Y, Hsientang Hsieh D P (2006) J Mass Spectrom 41: 169-174.

28. Alder L, Greulich K, Kempe G, Vieth B (2006) Mass Spectrom Rev 25: 838-865.

29. Gosetti F, Bolfi B, Manfredi M, Calabrese G, Marengo E (2015) J Sep Sci 18: 3130-3136.

30. Nürenberg G, Schulz M, Kunkel U, Ternes T A (2015) J Chromatogr A 1426: 77-90.

31. Swartz M E (2005) J Liq Chromatogr Relat Technol 28: 1253-1263.

32. Unger K K, Liapis A I (2012) J Sep Sci 35: 1201-1212.

33. Fekete S, Kohler I, Rudaz S, Guillarme D (2014) J J Pharm Biomed Anal 87: 105-119.

34. Guillarme D, Ruta J, Rudaz S, Veuthey J L (2010) Anal Biomed Chem 397: 1069-1082.

35. Fekete S, Veuthey J L, Guillarme D (2015) J Chromatrogr A 1408: 1-14.

36. Fekete S, Oláh E, Fekete J (2012) J Chromatrogr A 1228: 57-71.

37. Guiochon G, Gritti F (2011) J Chromatrogr A 1218: 1915-1938.

38. Deridder S, Desmet G (2011) J Chromatrogr A 1218: 46-56.

39. Gritti F, Guiochon G (2012) J Chromatrogr A 1252: 31-44.

40. Gritti F, Guiochon G (2012) J Chromatrogr A 1252: 45-55.

41. Gritti F, Guiochon G (2012) J Chromatrogr A 1252: 56-66.

42. Cabooter D, Fanigliulo A, Bellazzi G, Allieri B, Rottigni A, Desmet G (2010) J Chromatrogr A 1217: 7074-7081.

43. Gritti F, Cavazzini A, Marchetti N, Guiochon G (2007) J Chromatrogr A 1157: 289-303.

44. Fekete S, Ganzler K, Fekete J (2011) J Pharm Biomed Anal 54: 482-490.

45. Fekete S, Guillarme D (2013) J Chromatrogr A 1320: 86-95.

46. Gritti F, Guiochon G (2014) J Chromatrogr A 1333: 60-69.

47. González-Ruiz V, Olives A I, Martín M A (2015) Trends Anal Chem 64: 17-28.

48. Kostka J, Gritti F, Kaczmarski K, Guiochon G (2011) J Chromatrogr A 1218: 5449-5455.

49. Gritti F, Guiochon G (2010) J Chromatrogr A 1217: 1604-1615.

50. Gritti F, Tanaka N, Guiochon G (2012) J Chromatrogr A 1236: 28-41.

51. Gritti F, Guiochon G (2012) J Chromatrogr A 1238: 77-90.

52. Blackwell J A, Carr P W (1991) J Liq Chromatogr Relat Technol 14: 2875-2889.

53. Zhu C, Goodall D M, Wren S A C (2005) LC-GC N Am 23: 54-72.

54. Jackson P T, Carr P W (1998) Chem Tech 28: 29-37.

55. Fekete S, Fekete J (2011) J Chromatrogr A 1218: 5286-5291.

56. Iribane J V, Thomson B A (1976) J Chem Phys 64: 2287-2294.

57. Thomson B A, Iribane J V (1979) J Chem Phys 71: 4451-4463.

58. Dole M, Mach L L, Hines R L, Mobley R C, Ferguson L D, Alice M B (1968) J Chem Phys 49: 2240-2249.

59. Schmelzeisen-Redecker G, Butfering L, Rollgen F W (1989) Int J Mass Spectrom Ion Proc 90: 139-150.

60. Cech N, Enke C (2001) Mass Spectrom Rev 20: 362-387.

61. Wu J, Mc Allister H (2003) J Mass Spectrom 38: 1043-1053.

62. Arnhard C, Gottschall A, Pitterl F, Oberacher H (2015) Anal Bional Chem 407: 405-414.

63. Thurman E M, Ferrer I (2003) Liquid Chromatography/Mass Spectrometry, MS/MS and Time of Flight MS; Ferrer I, et al; ACS Symposium Series; American Chemical Society: Washington DC 14-30.

64. Stephens W E (1946) Phys Rev 69: 691.

65. Cameron A E, Eggers D F (1948) Rev Sci Instrum 19: 605-607.

66. Wiley W C, McLaren I H (1955) Rev Sci Instrum 26: 1150-1157.

67. Whittal R M, Li L (1995) Anal Chem 67: 1950-1954.

68. Brown S R, Lennon J J (1995) Anal Chem 67: 1998-2003.

69. Vestal M P, Juhasz P, Martin S A (1995) Rapid Commun Mass Spectrom 9: 1044-1050.

70. Cotter R J (1992) Anal Chem 64: 1027A-1039A.

71. Cotter R J (1999) Anal Chem 71: 445A-451A.

72. Mamyrin B A (2001) Int J Mass Spectrom Ion Proc 206: 251-266.

73. Morris H R, Paxton T, Dell A, Langhorne J, Berg M, Bordoli R S, Hoyes J, Bateman R H (1996) Rapid Commun Mass Spectrom 10: 889-896.

74. Petrovic M, Barceló D (2006) J Mass Spectrom 41: 1259-1267.

75. Kellmann M, Muenster H, Zomer P, Mol H (2009) J Am Soc Mass Spectrom 20: 1464-1476.

76. Kosjek T, Heath E, Petrovic M, Barceló D (2007) Trends Anal Chem 26: 1076-1085.

77. Hernández F, Pozo O J, Sancho J V, Lopez F J, Marin J M, Ibáñez M (2005) Trends Anal Chem 24: 596-612.

78. Krauss M, Singer H, Hollender J (2010) Anal Bioanal Chem 397: 943-951.

79. Wang N, Li L (2008) Anal Chem 80: 4696-4710.

80. Gosetti F, Bolfi B, Mazzucco E, Manfredi M, Marengo E (2015) in Advanced in Food Analysis research, chapter 10, Haynes Ed., Nova Publishers inc, USA.

81. Kuster B, Schirle M, Mallick P, Aebersold R (2005) Nat Rev Mol Cell Biol 6: 577-583.

82. Michalski A, Cox J, Mann M (2011) J Proteome Res 10: 1785-1793.

83. Bilbao A, Varesio E, Luban J, Strambio-De-Castillia C, Hopfgartner G, Müller M, Lisacek F (2015) Proteomics 15: 964-980.

84. Law K P, Lim Y P (2013) Expert Rev Proteomics 10: 551-566.

85. Hopfgartner G, Tonoli D, Varesio E (2012) Anal Bioanal Chem 402: 2587-2596.

86. Panchaud A, Scherl A, Shaffer S A, von Haller P D, Kulasekara H D, Miller S I, Goodlett D R (2009) Anal Chem 81: 6481-6488.

87. Weisbrod C R, Eng J K, Hoopmann M R, Baker T, Bruce J E (2012) J Proteome Res 11: 1621-1632.

88. Carvalho P C, Han X, Xu T, Cociorva D, Carvalho Mda G, Barbosa V C, Yates J R (2010) Bioinformatics 26: 847-848.

89. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R (2012) Mol Cell Proteomics 11: O111.016717.

90. Gosetti F, Bolfi B, Marengo E (2015) Anal Bioanal Chem 407: 4649-4659.

91. Gosetti F, Gianotti V, Polati S, Gennaro M C (2005) J Chromatogr A 1090: 107-115.

92. Gosetti F, Gianotti V, Mazzucco E, Polati S, Gennaro M C (2008) Dyes Pigm. 74: 424-432.

93. Gosetti F, Frascarolo P, Mazzucco E, Gianotti V, Bottaro M, Gennaro M C (2008) J Chromatogr A 1202: 58-63.

94. Gosetti F, Chiuminatto U, Mazzucco E, Calabrese G, Gennaro M C, Marengo E (2012) Anal Chim Acta 746: 84-89.

95. Gosetti F, Chiuminatto U, Mazzucco E, Calabrese G, Gennaro M C, Marengo E (2013) Food Chem 136: 617-623.

96. Gosetti F, Gianotti V, Angioi S, Polati S, Marengo E, Gennaro M C (2004) J Chromatogr A 1054: 379-387.

97. Gianotti V, Angioi S, Gosetti F, Marengo E, Gennaro M C (2005) J Liq Chromatogr Relat Technol 28: 923-937.

98. Gosetti F, Chiuminatto U, Mastroianni R, Mazzucco E, Manfredi M, Marengo E (2015) Food Additives and Contaminants A 32: 285-292.

99. www.massbank.jp

100. www.hmdb.ca

101. https://metlin.scripps.edu/index.php

102. www.nist.gov

103. Massart D L, Vandeginste B G M, Buydens L M C, De Jong S, Lewi P J, Smeyers-Verbeke J, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier Science B. V. 1997.

104. Miller J N, Miller J C, Statistics and Chemometrics for Analytical Chemistry, Prentice Hall 2000.

105. Martens H, Naes T, Multivariate calibration, John Wiley & Sons, London, 1989.

106. Kleinbaum D, Kupper L, Muller K, Applied regression analysis and other multivariate methods, 2nd ed. Pws-Kent, Boston, 1988.

107. De Noord O E (1994) Chemom Intell Lab Syst 25: 85-97.

108. Todeschini R, Introduzione alla chemiometria, EdiSES, Napoli, 1998.

109.   Douglass M Y, Cabezas H (1999) Comput Chem Eng 23: 1477-1491.

Chapter 2

Outline of the Thesis

The aim of this PhD thesis was to demonstrate the versatility of the LC-MS/MS technique in target and non-target analysis of cosmetic, food and environmental matrices; in particular, I considered a) the possible degradation of a dye used in nonpermanent tattoos, b) the possible variations in polyphenols content due to conservation time in extra virgin olive oil samples stored in different packaging materials and c) the possible degradation of three pesticides used in rice cultivations and the presence of other emerging contaminants that might be present in paddy water sample, thus evaluating water quality.

Electrospray ionization (ESI), both in positive and negative ion mode, was adopted as ion source. The identification of photodegradation compounds was based on the data provided by the quadrupole-time-of-flight mass spectrometry. All data, were acquired either using the Data Independent Acquisition (DIA) and the Data Dependent Acquisition (DDA) mode.

A direct comparison between the investigated sample and the control was always performed in order to provide a first screening of the acquired data; such comparison provide the enlightenment of only the significant *m/z* signals related to non-target compounds that might be ascribable both to compounds naturally present within the investigated matrix and to new species formed therein along time. Furthermore, to reduce the redundant information, extracting only the useful one, chemometric methods (Principal Component Analyses, PCA, Partial Least Square, PLS, and Discriminant Analysis, DA) were employed. Ions of interest were further processed using both online database (i.e. ChemSpider) and in-house libraries, searching for correspondences between the MS spectrum and the

hypothesized structures. The proposed structures were automatically compared to the recorded accurate MS/MS spectra on the basis of knowledge of breaking and rearranging bonds. As a result, a matching factor assessing the goodness of the matching between the experimental pattern and the theoretical fragmentation using the proposed structure was calculated. If the matching percentage was not satisfactory, it was necessary to interpret the acquired spectra in order to attribute a chemical structure to the unknowns.

The complementarity of the information provided by the target and non-target approaches demonstrated to be fundamental for the complete understanding of the matrices object of study. Furthermore, the digitalization of each sample, provided by the chosen unselective DIA acquisition mode, will enable future re-interrogation of the same sample (retrospective analysis) without the need of further analysis: this could be useful if there will be the need to evaluate the presence of new molecules considered hazardous, whose concentration will be strictly regulated by new normative.

Chapter 3

Identification of sulforhodamine B photodegradation products present in nonpermanent tattoos by micro liquid chromatography coupled with tandem high-resolution mass spectrometry

3.1 Introduction

Dyes are additive substances able to produce chromatic variations to substrates to which they are added. The physical effect of the dye is due to the absorption and transmittance of a specific wavelength of the visible light. It is possible to describe a dye as a material constituted by two different parts: a *chromogen* that acts as an electron acceptor and an *auxochrome* that plays the role of an electron donor and makes the dye soluble and dyeing. The first part consists of an aromatic body linked to the group responsible of the color: together they constitute the *chromophore* group. As it absorbs light in the visible region, it generates shifts (blue shift, *hypsochromic*, or red shift, *bathochromic*) of the wavelengths in UV-Vis region, causing a color change.

It is possible to classify the dyes in three different categories: colorants, pigments and lakes. Organic colorants are soluble or partially soluble in water, giving color through inclusion processes or chemical reactions. Pigments, instead, are powder and, as they are water insoluble, they have to be suspended into a ligand media that can fix them to the substrate. Finally, lakes are particular dyes that are precipitated with a *mordant*, or inert binder, typically a metallic salt.

Some dyes are used to decorate body in the art of tattoo. This represents an ancient form of art, firstly employed in prehistory, spread in every part of the world and become more and more popular in recent years [1-9]. Different reasons induce people to be tattooed: the most popular are related to group or tribal affiliation, social identification and personal artistic expression. Permanent tattoos consist in the application of inks constituted by suspended colored crystals of about 0.1-10 µm [10] in the dermis, trough a metallic needle. In order to give brightness to the shades, titanium dioxide is widely used [11]. Only a part of the applied ink remains in the dermis, coloring the picture: the lymphatic system removes the remaining part from the skin since macrophages identify pigments as foreign bodies [12-14].

In the literature there are many studies about the permanent ink toxicity, but they are mainly focused on metal allergies (nickel, lead and cadmium) and their degradation pathways [15-18]. Sunlight [15,19] and laser light [16] are the most studied degradation agents since they respectively represent natural condition of dermal exposure and the practice of laser surgery tattoo removal. 2-methyl-5-nitraniline, 4-nitrotoluene, 2,5-dichloroaniline and 1,4 dichlorobenzene were found as degradation products, toxic for humans. A study performed by Cui et *al*. in 2004 [15] simulated the action of UV-A and UV-B radiation (21:1 ratio) on Pigment Yellow 74. The obtained results demonstrated that the pigment undergoes photolysis under solar light or more intense irradiation sources, leading to the formation of toxic photodegradation products, namely *N*,*N*-bis(2-methoxyphenyl)urea, 2-(hydroxyimine)-*N*-(2-methoxyphenyl)-3-oxobutanamide and *N*-(2-methoxyphenyl)-3-oxobutanamide.

Laser light effects were studied on two monoazo dyes CI 12315 and CI 12460 by Vasold et *al*. [16], with the aim of simulating conditions of tattoo removal, evidencing the formation of 2-methyl-5-nitroaniline, 2,5-dichloroaniline, 1,4-dichlorobenzene and 4-nitrotoluene, all classified as carcinogenic degradation products.

The same degradation products, with the addition of naphtol AS and methoxynaphtol AS, were evidenced from the photodegradation of Pigment Red 22 (CI 12315) by Engel et *al.* [19]. The pigment was dissolved in tetrahydrofuran and then added to dioxane, dichloromethane, chloroform, and then subjected to UV lamp irradiation (280-320 nm, I=0.0015 W/cm^2). The degradation products were toxic on inhalation and on contact with dermis.

Nonpermanent tattoos represent an alternative of the classic ones, if body decoration is desired. They are common not only in children, but also in adults that want to temporarily decorate their bodies. Nonpermanent tattoos are common items since they can be bought individually or included as gadgets with sweets, periodicals and comics. The most popular nonpermanent tattoos are made with henna, an extract of *Lawsonia inermis* [20], while others are based on xanthene, diarylheptanoid or azo dyes [21]. Recently, growing concern on collateral effects of nonpermanent tattoos has increased, driven by their possibility of causing allergies [22]. It has to be underlined that the literature on this topic is mainly focused on the toxicity of the dye itself when it is used for tattooing, but not on their fate once exposed to stressing sources [23, 24]. Eleven dyes used in nonpermanent tattoos were object of the study conducted by Madan and Beck [23]: all the considered dyes showed allergic features and, furthermore, one of them (CI 15850) was present at too high concentration, resulting to be even more hazardous for health.

The allergic features of dyes used in cosmetics were recently investigated by different studies [24], some of them were based on quantitative structure-activity relationship (QSAR) models. Such approaches demonstrated how CI 11920 could be classified as a medium to strong allergen, whereas CI 45100 as a weak one.

At our knowledge, in the literature there are no studies focused on the possible degradation of nonpermanent tattoo dyes due to sunlight and the identification of their photodegradation products. It is well known that atoxic permitted dyes can be degraded by solar radiations and lead to toxic compounds [19, 25-30].

For this reason, the research work presented in this chapter is focused on the study of sulforhodamine B, also called Acid Red 52 (AR52), a permitted dye largely used in non-permanent tattoos. AR52 is constituted by a xanthenic core structure linked to a disulfonated benzene ring (Fig. 3.1).
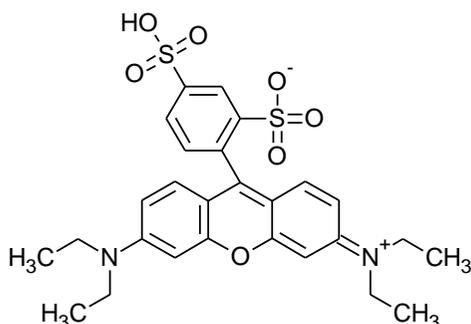
Figure3.1: chemical structure of sulforhodamine B, also called Acid Red 52 (AR52).

Since the Magnuson-Kligman test performed on AR52 by the Scientific Committee on Cosmetic Products and Non-Food Products [31] gave negative results, the dye was classified as a nonsensitizier. Conversely, the QSAR approach confuted the Magnuson-Kligman test since the results showed a weak allergen attitude, hence the dye must be applied on the skin just for short periods [32].

The aim of this part of the thesis is the investigation of AR52 and the identification of its possible degradation products when applied on skin and exposed to sunlight irradiation [33]. For this aim, simpler AR52 model systems were studied first (ultrapure water solution and the sweat-simulating one) before considering the more complex model constituted by pigskin.

Since summer is the period in which nonpermanent tattoos are mostly applied on skin, the selected degradation conditions aimed to simulate that period.

The analysis were performed using micro liquid chromatography (LC) coupled with tandem high-resolution mass spectrometry (MS), working in data dependent mode. As there was no *a priori* knowledge about the possible photodegradation species that could be generated, it was mandatory to collect all possible information

about the unknown species through the triggering of more MS experiments, i.e. full-scan MS and high-resolution MS/MS. The obtained results were evaluated using chemometric approaches (principal component analysis, PCA, and discriminant analysis, DA) in order to reduce the redundant information, extracting only the useful one such as the *m/z* signals of the newly formed degradation products.

3.2 Experimental

3.2.1 Reagents

Reagent	Purity	Vendor
Methanol	LC-MS Ultra CHROMASOLV, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Ammonium formate	> 99.995 %	Sigma–Aldrich (Milwaukee, WI, USA)
Sodium chloride	99% or greater	Sigma–Aldrich (Milwaukee, WI, USA)
Acetic acid	Eluent additive, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Ammonia	30 %	Sigma–Aldrich (Milwaukee, WI, USA)
Sodium hydroxide	98 % or greater	Sigma–Aldrich (Milwaukee, WI, USA)
AR52	75 %*	Sigma–Aldrich (Milwaukee, WI, USA)
Water	For mass spectrometry	Fluka (Buchs, Switzerland)
L-histidine	> 99.5 %	Fluka (Buchs, Switzerland)
Sodium phosphate monobasic	> 98 %	Fluka (Buchs, Switzerland)

Table 3.1: list of the reagents used within the research project.

* AR52 75 % dye content was chosen to reproduce real degradation conditions since the dye is commercially available in cheap items where the dye is not used in high purity degree.

3.2.2 Apparatus

In order to reproduce natural sunlight, a Solarbox 3000e (CO.FO.ME.GRA, Milan, Italy) was used to provide continuous availability and constant intensity. The irradiation intensity of the xenon lamp (λ=280-800 nm) and temperature were respectively set at 800 Wm^{-2} and 35°C, in order to simulate the period from June to August in Italy, in which nonpermanent tattoos are mainly applied. The selected values were chosen on the basis of the meteorological conditions typical of the city of Alessandria (Piedmont, Italy), monitored by the weather station of our Department.

The ultra-high-performance LC-MS/MS analysis were performed with a micro LC system (Eskigent Technologies, Dublin, CA, USA) that included a micro LC200 pump with a flow module of 5-50 μL and CTC programmable array logic autosampler with a Peltier unit (1-45 °C). The system was hyphenated with a 5600$^+$ Triple TOF^{TM} system (Sciex, Concord, Canada) equipped with a DuoSpray^{TM} ion source and calibrant delivery system. The 5600$^+$ Triple TOF^{TM} data were processed by PeakView^{TM} 2.1 (Sciex, Concord, Canada) in order to investigate and interpret accurate mass spectral data and by MasterView^{TM} 1.0 (Sciex, Concord, Canada) in order to compare the intensity of mass spectral data among different samples. MarkerView^{TM} 1.2 (Sciex, Concord, Canada) was used to analyze the acquired data and to provide graphical representations.

Software T.E.S.T. version 4.1 (US Environmental Protection Agency) was adopted to perform the virtual mutagenicity test.

3.2.3 Stock solutions and sample preparation

The AR52 stock standard solution was prepared in ultrapure water at 100.0 mg L^{-1}, diluted as required in ultrapure water and stored at 4 °C in dark glass vials. The 100.0 mg L^{-1} solution simulating sweat was prepared according to Piunti et *al.* [34]: 0.2 g L-histidine, 2 g sodium chloride, 0.88 g sodium phosphate monobasic and 3.15 mL hydrochloric acid (1M).

The pretreatment of the real samples consisted in the application of 40 μL of 50.0 mg L^{-1} aqueous AR52 solution on pigskin (2 x 3 cm), letting the aqueous solution to evaporate for 12 minutes. Successively, 20 μL of the dye were deposited on the skin and after 12 minutes, the last 10 μL were applied. Three genuine replicates were obtained through the application of the described deposition procedure to other two pieces of pigskin. The samples were put into the Solarbox and irradiated. Ultraturrax was used to homogenize the irradiated samples with a mixture of water and acetonitrile (50:50 v/v). The samples were centrifuged (4 °C, 7000 rpm, 7 min), filtered twice, first through a 0.45-μm PTFE filter (VWR International, Darmstadt, Germany) and successively through a 0.20 μm PTFE filter (VWR International, Darmstadt, Germany). The prepared samples were stored in dark vials at -20 °C until the analysis.

In order to evaluate the photoirradiation effects on the whole solution and to consider all the possible interactions that could occur between AR52 and the compounds present in the sweat-simulating solution and in the pigskin, the samples were not subjected to any purification step.

3.2.4 Chromatographic and mass spectrometric conditions

Halo C18 column (0.5 mm x 100 mm, 2.7 μm; Eksigent Technologies, Dublin, CA, USA) was selected as the stationary phase. The mobile phase was a mixture of water (solvent A) and methanol (solvent B) both with the addition of 0.1 % (v/v)

glacial acetic acid, eluting at a flow rate of 40 µL min^{-1} in the gradient reported (Tab. 3.2):

Time (min)	% A	% B
0.00	90.0	10.0
0.15	90.0	10.0
5.00	5.0	95.5
5.50	5.0	95.5
5.60	90.0	10.0
8.00	90.0	10.0

Table 3.2: elution gradient.

Injection volume was set at 1 µL , while the column oven temperature was 45 °C. The DuoSpray^{TM} ion source worked both in positive ion mode (PI) and in negative one (NI). The instrumental conditions were set as follows: curtain gas (N$_2$) at 25 psig, nebulizer gas GS1 and GS2 at 25 and 35 psig respectively, desolvation temperature at 350 °C, collision activated dissociation gas at 5 units of arbitrary scale of the instrument, and ion spray floating voltage at 5300 or -4300 V depending on the selected polarity. During time-of-flight (TOF) MS experiments, the declustering potential and the collision energy were respectively set at ±100 and ±10 V, while during the TOF-MS/MS experiments the collision energy was set at ±70 V, depending on the adopted polarity.

The software tool Information Dependent Acquisition triggered a TOF MS scan (100-1000 amu, accumulation time 0.30 s) as a "survey scan" and up to five dependent TOF-MS/MS scans (50-1000 amu, accumulation time 0.10 s) in order to obtain as much information as possible about the compounds that could be generated during photodegradation. The survey scan and the dependent scans were sequentially performed for the whole duration of the chromatographic run (cycle

time of 0.85 s), in order to obtain the maximum information from a single micro LC-MS/MS analysis, i.e. the highly accurate and sensitive full-scan MS and MS/MS acquired spectra.

3.3 Results and discussion

3.3.1 Photodegradation experiments on the AR52 standard aqueous solution

Fourteen milliliters of AR52 aqueous solution (50.0 mg L^{-1}) were subjected to photoirradiation using Solarbox. The pH of the solution was adjusted to 5.5 with HCl in order to have the same value of the sweat-simulating solutions and to avoid possible pH effects on dye degradation. The solution was introduced in a quartz cell and subjected to Solarbox irradiation for 9 days, as described in Paragraph 3.2. The brightness of the AR52 solution progressively vanished as a function of the irradiation time. Whenever a color change occurred, an aliquot of the solution was collected (in total 14 aliquots) and then subjected to micro LC-MS/MS analysis.

The degradation reactions are due to the effect of photoirradation and not to temperature alone, as demonstrated by previous experiments that tested the temperature effects on the dye solutions up to 60 °C that did not show any color change. Plotting the ln A (A corresponds to the area of AR52 chromatographic peak) as a function of the degradation time, a linear trend was observed, suggesting that AR52 photodegradation follows a first order kinetic.

The calculated half-life is about 1.9 days.

3.3.2 Photodegradation experiments on the AR52 sweat-simulating solution

The experiments performed on model aqueous solutions were performed also on the sweat-simulating ones. Fourteen aliquots of the solution were collected at different irradiation times and subjected to micro LC-MS/MS analysis. the

degradation rate demonstrated to be slightly faster into the sweat-simulating solution than in the aqueous one, indicating an accelerating effect stimulated by other components of the solution on the degradation process. Once again, plotting ln *A* (*A* represents the chromatographic area of AR52 peak) as a function of time, a linear behavior was evidenced, thus also in this matrix photodegradation process follows a first-order reaction. The calculated half-life is about 1.6 days.

The difference occurring between the two kinetics could not be attributed to pH effect since the aqueous solution pH value was corrected to 5.5, in order to have the same value of the sweat-simulating one.

3.3.3 Development of the LC-MS/MS method

In order to obtain useful information on the possible degradation products generated by the irradiation of the dye characterized by the presence of both sulfonate and amino groups (Fig. 3.1), AR52 was characterized using both negative (NI) and positive one (PI) ionization mode, as it is possible that some product ions generated during the MS/MS experiments could have a structure compatible with that of the degradation products of the dye. For this aim, 100 μg mL^{-1} of AR52 solution in methanol-acetic acid (99.9/0.1 v/v) were infused into the mass spectrometer through a syringe pump (flow 0.1 μL min^{-1}). The most intense acquired signal for AR52 in NI was at *m/z* 557.1426, corresponding to the [M-H]$^{-}$ species, while in PI the most intense signal was at *m/z* 559.1560 corresponding to the [M+H]$^{+}$ species (Fig. 3.2 and 3.3).
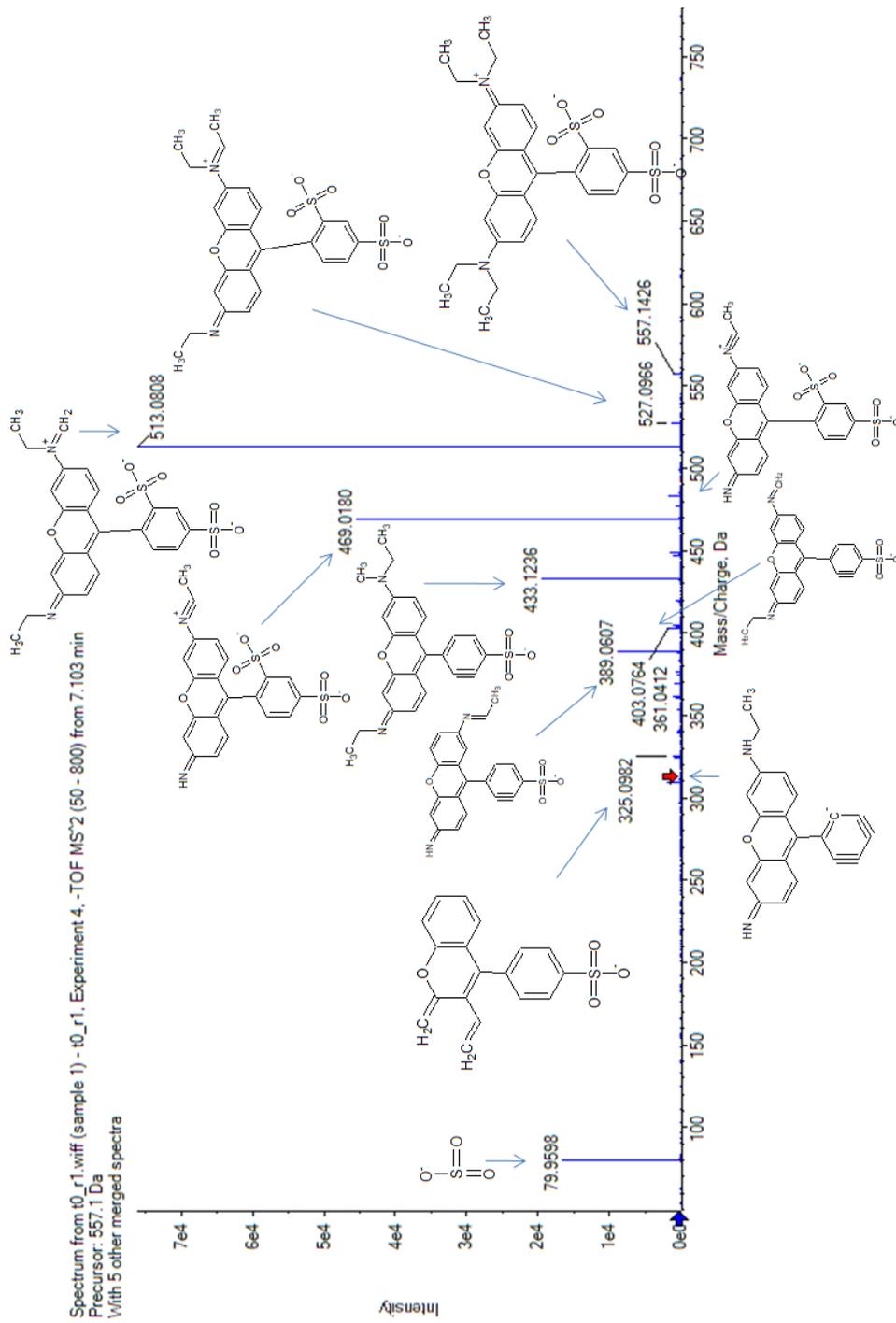
Figure 3.2: AR52 product ion spectrum (MS/MS) of *m/z* 557.1426 (NI) and the chemical structures of the most abundant signals.
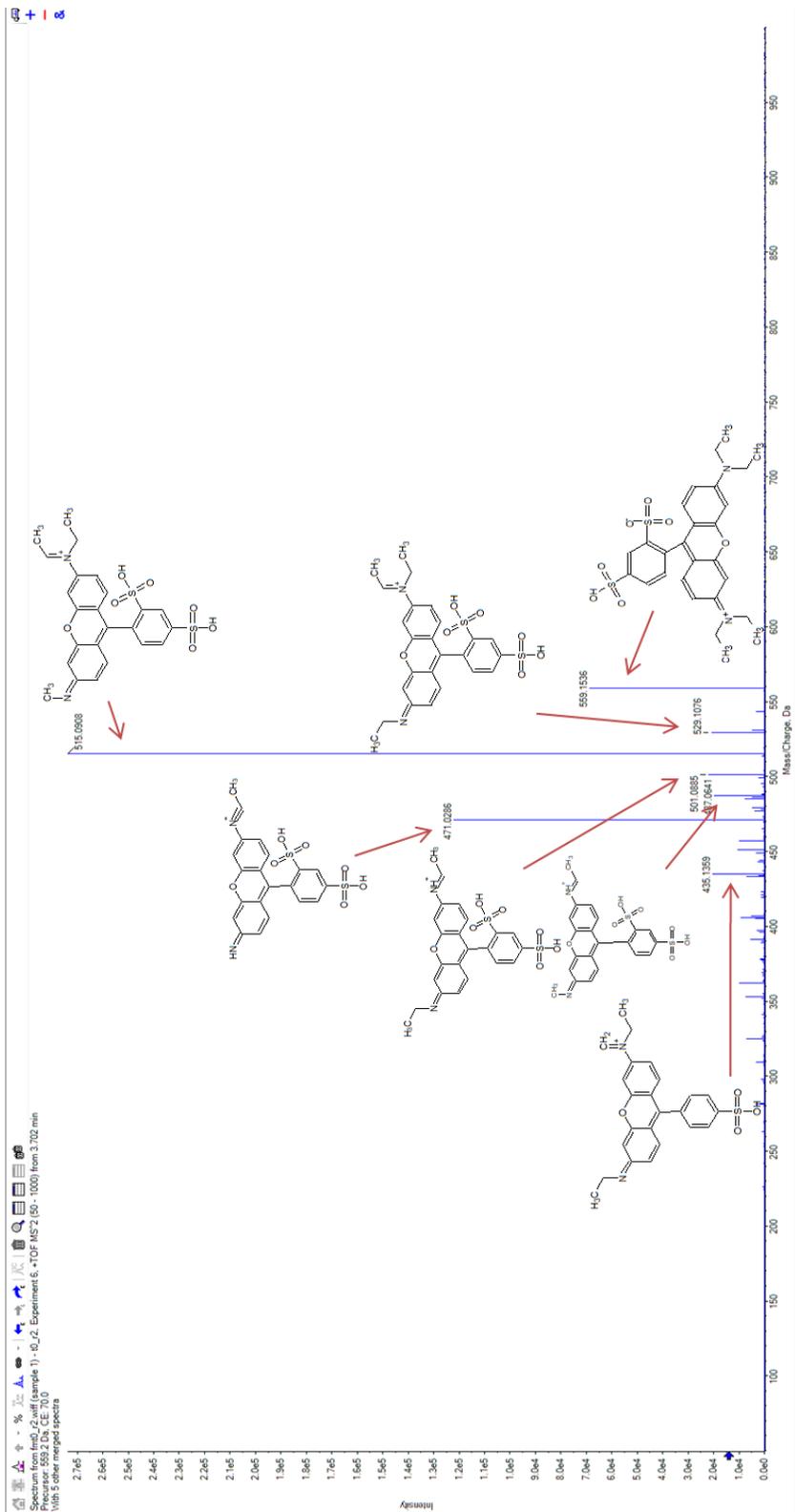
Figure 3.3: AR52 product ion spectrum (MS/MS) of at *m/z* 559.1560 (PI) and the chemical structures of the most abundant signals.

For the chromatographic analysis, a Halo C_{18} column was selected as the stationary phase, as it is able to provide good retention of xanthenic compounds, such as AR52. Acetic acid was added to the mobile phase in order to improve not only the dye ionization in both modalities, but also its chromatographic peak shape.

A calibration plot was built at seven concentration levels: 10.0 µg L^{-1}; 25.0 µg L^{-1}; 50.0 µg L^{-1}; 100.0 µg L^{-1}; 250.0 µg L^{-1}; 500.0 µg L^{-1}; 1000.0 µg L^{-1} ; (three replicates for each level) that reports the *m/z* 557.1426 peak area as a function of its concentration. The calibration line was weighted by a 1/x factor, the equation of the line was y=114.06x-24 and its R^2 value was 0.9985. The Limit Of Detection (LOD) and Limit Of Quantitation (LOQ) were expressed as analyte concentrations. The former represents the minimum detectable quantity of the analyte in a matrix and it is evaluated as the concentration that provides a signal that is at least greater than three times the standard deviation of the noise signal. The latter represents the minimum quantifiable amount of the analyte within the matrix. LOQ is about ten times the standard deviation of the noise signal. LOD and LOQ obtained values were 1.7 µg L^{-1} and 5.8 µg L^{-1} respectively.

Intra- and inter-day reproducibility were also evaluated on both dye retention times and chromatographic peak areas, analyzing the dye (1.0 mg L^{-1}) every day (five replicates) along a week (Tab. 3.3).

	AR52 standard aqueous solution		AR52 sweat-simulating solution	
	RSD% Intraday	*RSD% Interday*	*RSD% Intraday*	*RSD% Interday*
t_R	0.9	1.5	0.19	0.51
Area	3.80	5.01	1.69	5.46

Table 3.3: RSD% calculated on the retention times and peak areas of the dye.

3.3.4 Strategies for the identification of the unknown species

Total ion chromatograms of AR52 aqueous solution (5.0 mg L^{-1}) before and after 226 hours (9.42 days) of photoirradiation are reported in Figure 3.4.

AR52 peak intensity (t_R= 3.79 min) decreases as the degradation occurs, up to totally disappear. Apparently, no other peaks corresponding to degradation products afterward indicated as P1, P2 etc., were evidenced. On the contrary, it was possible to see chromatographic peaks present before photoirradiation. These peaks are probably ascribable to the impurities present in the AR52 standard, since its purity is only 75%.

For the identification of the unknown species a non-target approach was adopted, using both direct comparison and Principal Component Analysis as described in Paragraph 1.4.6.
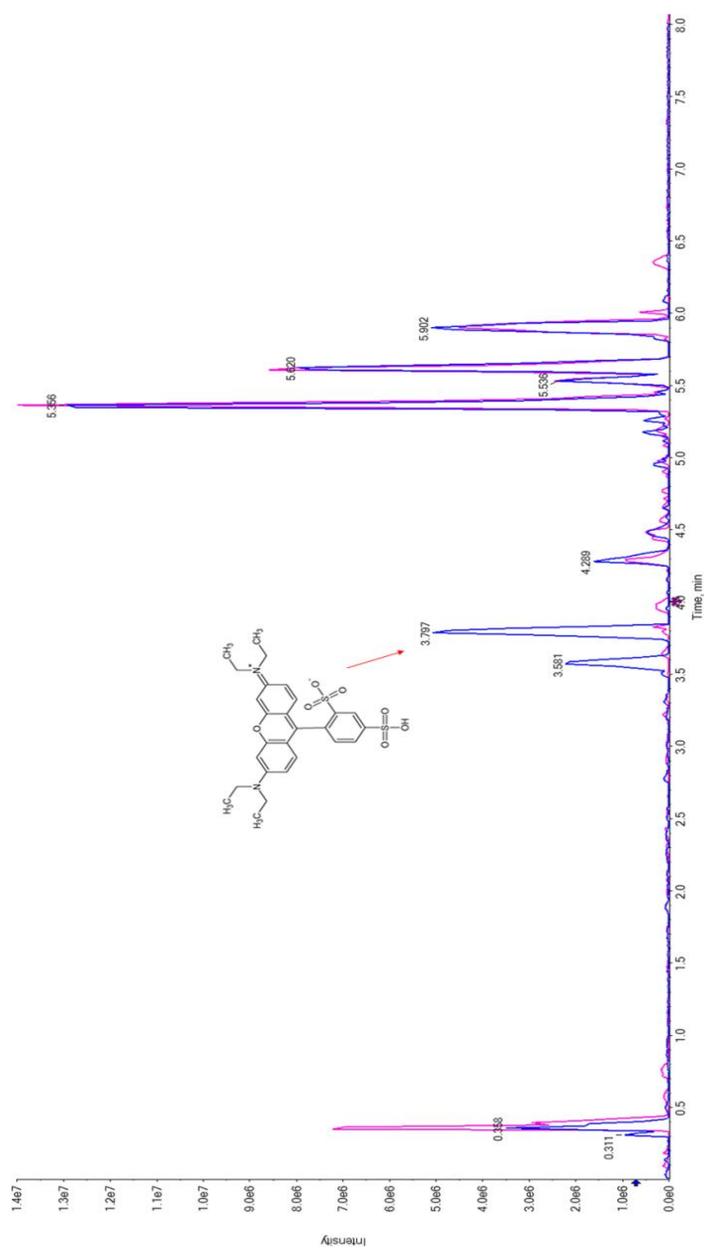
Figure 3.4: total ion chromatogram of AR52 aqueous solution before the degradation (*blue trace*) and after 226 hours of photoirradiation (*magenta trace*).

A "*direct comparison*" between the non-irradiated samples (control) and those that were subjected to photoirradiation was performed to underline if the impurities present in the standard dye undergo degradation. The chromatogram of a non-irradiated sample (control sample) was compared with that of an irradiated sample (degraded sample) through Peak-View^{TM} software, after setting the following parameters:

- Do not calculate details for XIC with intensity <5000 counts;
- Signal to noise ratio <500;
- XIC width (Da) = 0.02;
- Retention time width (min) = 2;
- Threshold (cps) = 1000;
- Threshold (ratio of control) = 100;
- Minimum retention time = 0.25 min;
- Maximum retention time = 5.5 min.

Through such comparison it was possible to highlight the differences occurring between the two considered chromatograms. The degraded peaks are considered significant by the software only if they are 100 times more intense than the corresponding peaks in the control sample. Automatically, all the significant signals were recognized by the software, included in a list, and successively extracted from the total ion current (TOF scan). The new species formed during the photodegradation process were evidenced and their structures were hypothesized. The high chromatographic reproducibility, essential to provide reliable results in direct comparisons was evaluated through the intra- and inter-day per cent precision, respectively below 0.9% and 1.5%, as reported in Paragraph 3.3.3.

The setting of these criteria allowed the identification of the two main impurities present in the standard dye, probably originated from the industrial synthesis of the

dye itself. They were present before the irradiation (t_0) and their intensity decreased up to vanish at the end of the irradiation process (t_f), following the same kinetic of AR52 (Fig. 3.5). This aspect had to be evaluated in the elucidation of the AR52 degradation products' structures, since the new compounds could be generated not only from the dye, but also from its impurities. Impurities *I1* and *I2* (retention times at 3.31 and 3.58 min, respectively) are characterized by *m/z* 503.0926 and *m/z* 531.1236 in PI, respectively.
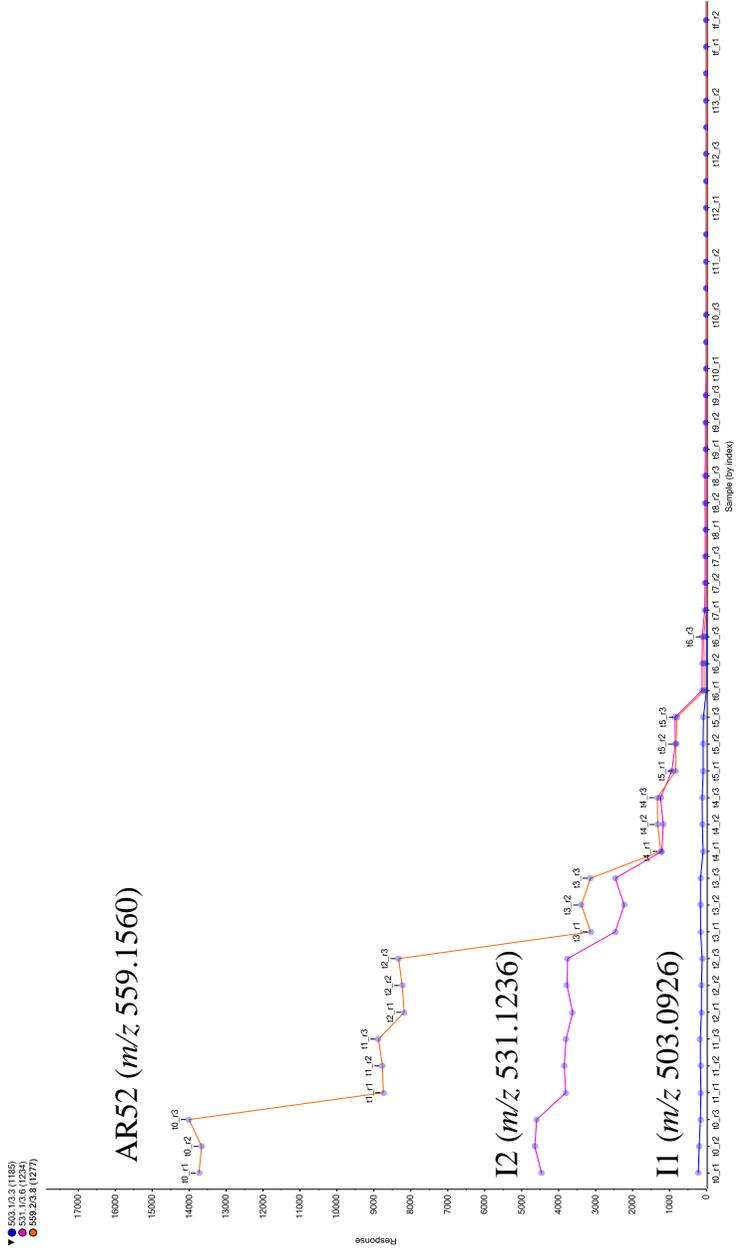
Figure 3.5: *m/z* signal intensities of AR52 (*orange trace*), *I1* (*blue trace*) and *I2* (*magenta trace*) as a function of the irradiation time.

Given the great amount of the acquired data (MS and MS/MS spectra of a large number of samples), the multivariate approach was adopted for the identification of the degradation products, since it compares all the samples with each other in a unique step, and not only two at a time as it occurs in the case of direct comparison. Principal Component Analysis (PCA) was used for this aim. In order to select the variables (i.e. the *m/z* values obtained for all the acquired chromatograms), an alignment algorithm was applied to all the chromatographic peaks present in all the different chromatograms. The parameters used to extract the significant peaks from the raw data were set as follows:

- Minimum retention time to ignore the void volume = 0.25 min;
- Maximum retention time (the final time of the chromatographic run before re-equilibration of the column) = 5.5 min;
- Subtraction offset = 5 scans;
- Subtraction multiplication factor = 2;
- Noise threshold = 200;
- Minimum spectral peak width = 20 ppm;
- Minimum retention time peak width = 5 scans;
- Retention time tolerance = 0.30 min;
- Mass tolerance = 10.0 ppm;
- Maximum number of peaks = 5000.

Such constraints allowed the software to find a small number of *m/z* signals that were merged during the alignment process.

The final dataset consisted of 90 samples (1 AR52 aqueous solution before degradation, 1 AR52 sweat-simulating solution before degradation, 14 AR52 aqueous solutions degraded at different irradiation times, 14 AR52 sweat-simulating solutions degraded at different irradiation times; each

irradiation/analysis was replicated three times) and 1735 variables corresponding to the different *m/z* values obtained from the alignment of the 90 chromatographic runs. In order to eliminate possible scale effects and to make the variables to account for the same amount of information, the data were scaled before performing PCA using the Pareto scaling [35].

The score (on the left) and the loading (on the right) plots of the first two PCs are reported in Figure 3.6. The high correlation occurring between the data is expressed by the sum of the percentages of the information explained by the first two PCs, 28.5% and 21.7 % for PC1 and PC2 respectively, representing more than 50% of the overall variance. It can be observed that the data are clearly separated in four groups. AR52 sweat solution samples before irradiation are at the right bottom (green ellipse), whereas the degraded samples in the same matrix are at the left bottom (red ellipse). Given the non-irradiated AR52 aqueous solutions (t_0) in the blue ellipse close to PC1 axis, it is possible to identify their degradation as a function of the irradiation time (from t_1 to t_7) following the arrow up to the left upper zone of the plot.
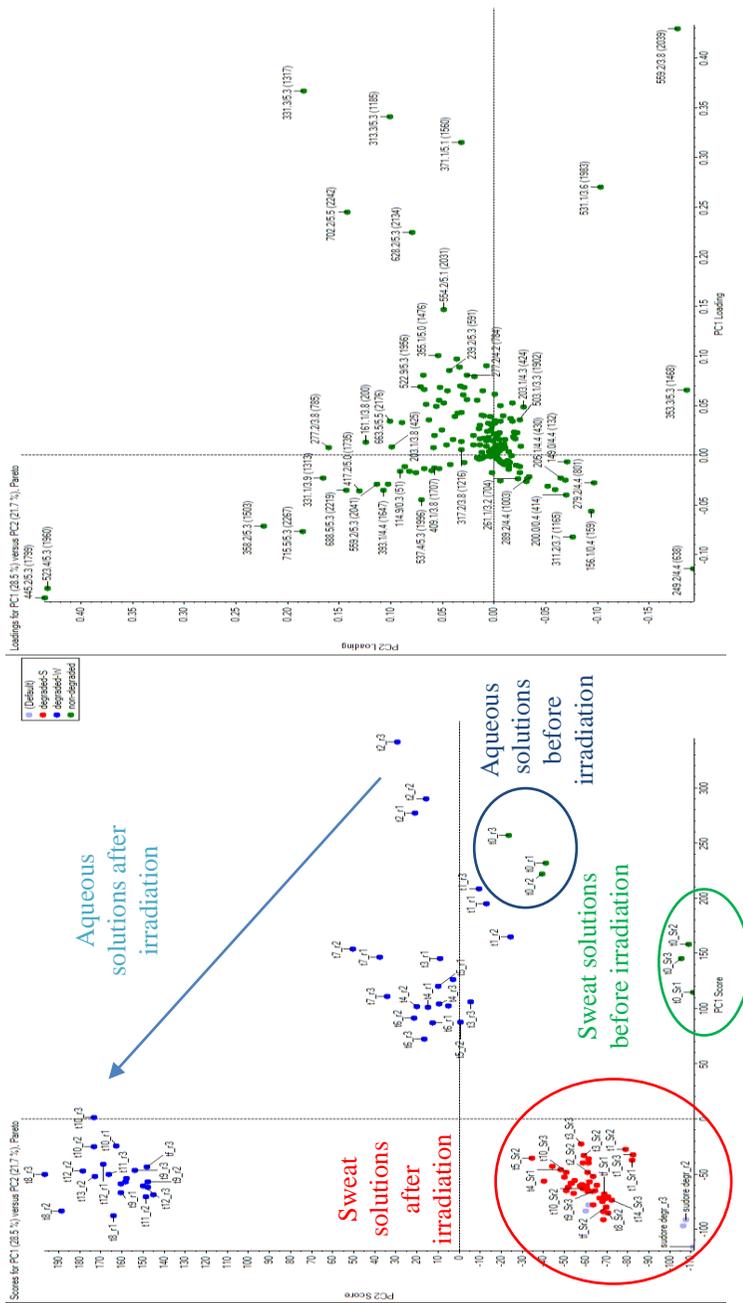
Figure 3.6: score (*left*) and loading (*right*) plots of the first two Principal Components (*PC*). The ellipses indicate the different groups of the analyzed samples.

103

Observing the PC1 axis, from positive to negative values it is possible to distinguish the non-degraded from the degraded samples, while looking at the PC2 axis, from positive to negative values it is possible to notice the separation between the two analyzed matrices: the aqueous solutions and the sweat-simulating solutions. It follows that the degraded solutions are characterized by negative loadings (*m/z* values) on PC1, and in particular the aqueous solution group has positive values on PC2, while the sweat-simulating solution group has negative values on PC2.

3.3.5 Elucidation of the degradation product structures

The chemical structures of the new species formed during the degradation processes were proposed evaluating the elemental composition within a mass tolerance of 3.6 ppm, the abundance of the isotopic cluster, the number of ring and double bonds, the accurate MS/MS spectrum to obtain the elemental composition of the product ions (average mass tolerance less than 6.3 ppm) and lastly the reasonability of the retention time considering the correlation with the calculated Log*P* of the unknown species. The interpretation of the high-resolution and accurate product ion MS/MS spectrum played a key role in the assignment of putative structures of the precursor ions. Once each chemical structure had been proposed, it was drawn in Peak-View[TM], where its MS/MS spectrum was simulated *in silico* and compared to the experimental spectrum, giving a matching percentage. A significant positive match was obtained for all the hypothesized molecular structures.

For all the degradation products shown in Figure 3.7, the accurate mass, the elemental composition, the number of rings and double bonds, the mass error (ppm), and the average MS/MS product ions are reported in Table 3.4. All the proposed structures shown in Figure 3.7 are characterized by the conservation of

the sulfonate benzene skeleton of the dye, the opening of the xanthenic core and the loss of one or both the linear carbon chains linked to the nitrogens.

$[M+H]^+$ species	Elemental composition	Accurate m/z	t_R (min)	RDB	MS (ppm)	MS/MS (ppm)
AR52	$C_{27}H_{30}N_2O_7S_2$	559.1560	3.79	14	3.6	4.5 (6)
P1	$C_{19}H_{20}O_7S$	393.1002	4.43	10	-3.2	3.6 (4)
P2	$C_{21}H_{33}NO_5S$	412.2152	5.47	6	-3.2	4.5 (8)
P3	$C_{22}H_{36}O_7S$	445.2258	5.30	5	0.6	12.0 (17)
P4	$C_{22}H_{28}O_8S_2$	485.1294	5.47	9	-0.3	11.5 (57)
P5	$C_{40}H_{66}N_4O_5S$	715.4825	5.30	10	1.0	9.2 (13)
I1	$C_{23}H_{22}N_2O_7S_2$	503.0926	3.31	14	-3.4	-3.4 (8)
I2	$C_{25}H_{26}N_2O_7S_2$	531.1236	3.58	14	0.5	1.8 (6)

Table 3.4: elemental composition, accurate *m/z*, retention time, number of rings and double bonds (*RDB*), mass error, and intensity-weighted average error for the tandem mass spectrometry (MS/MS) data calculated through the number of the product ions reported in *parenthesis*.
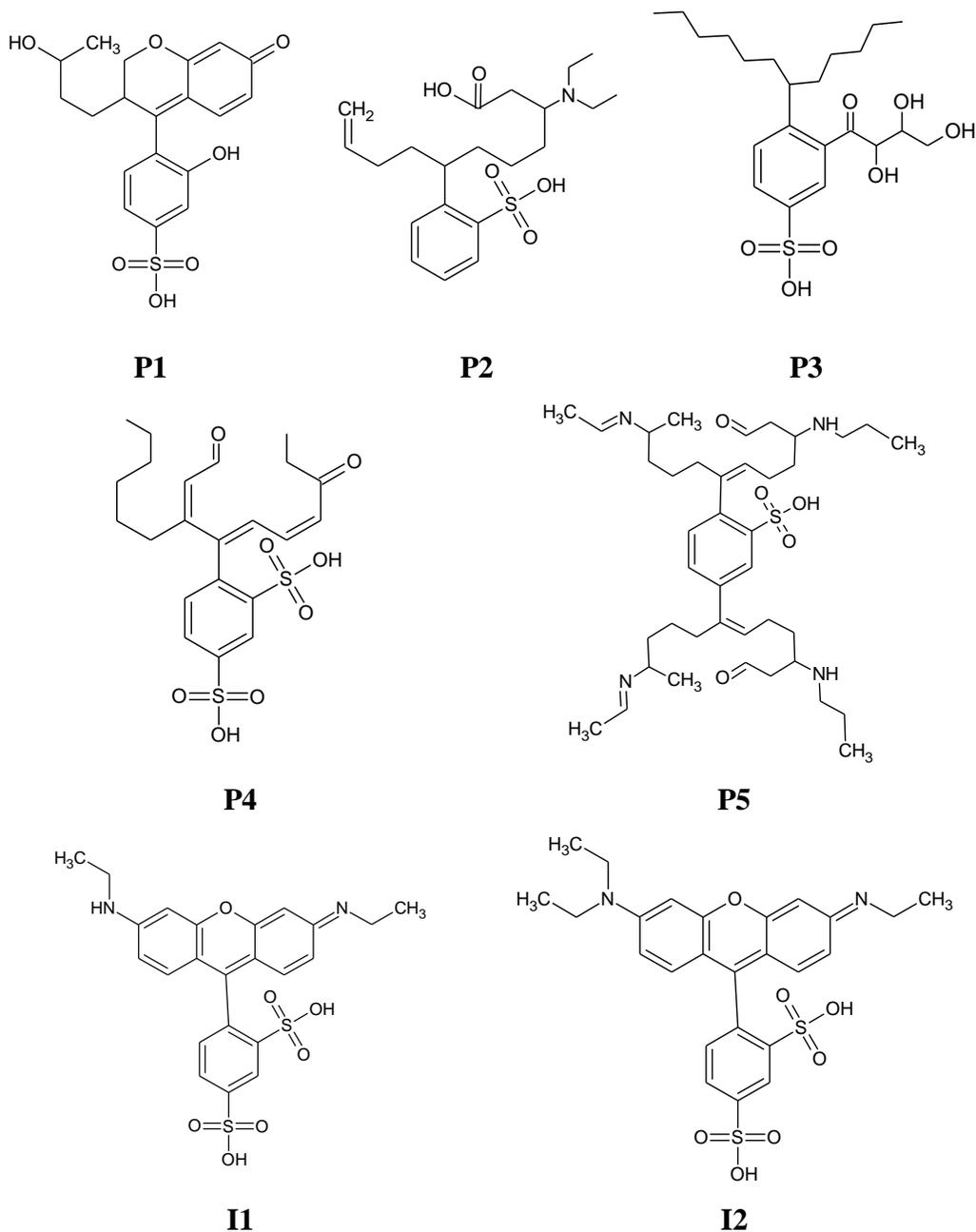
Figure 3.7: putative chemical structures of the AR52 photodegradation products and impurities.

On the basis of the literature focused on rhodamine B [36, 37], we hypothesized both the irradiation and the presence of oxygen as AR52 degrading agents [38]. After UV-Vis light irradiation, AR52 is promoted to the excited state, able to

reduce O_2 to $O_2^{\bullet-}$. The newly formed species turns into $OOH^{\bullet}$ after the reaction with a proton from autoprotolysis of water (solvent). $OH^{\bullet}$ and $OOH^{\bullet}$ attack AR52 inducing its N-deethylation, essential for the opening of the xanthenic core and the consequent color fading. A second theory could be suggested considering the simultaneous N-deethylation and breaking of the chromophore structure of the dye as competitive mechanisms occurring during photoirradiation [39-41]. The competition between the two mechanisms is due to the formation of the radical species: the formation of a nitrogen-centered radical leads to the N-deethylation mechanism, whereas the formation of a carbon-centered radical leads to the breaking of the dye chromophore structure [40, 41]. Carbon dioxide, mineral acids and water are not the final results of the AR52 photodegradation process, as it happens for rhodamine B, because its two sulfonic groups act as protective agents as reported for other sulphonated molecules [42].

A common degradation pathway had been suggested for both the AR52 tested media since there were no differences occurring between the degradation products identified in aqueous and sweat-simulating solutions and the respective degradation rates were only slightly different.

3.3.6 Degradation of tattooed pigskin and identification of the degradation products

To better simulate what happens with AR52 tattooed skin, the dye was applied on pigskin as described in Paragraph 3.2.3. Pigskin and tattooed pigskin (three replicates for each sample) were subjected to photoirradiation using the same conditions adopted for the AR52 solutions degradation. The discoloring of the dye applied on the pigskin was completed in only 55 h (Fig. 3.8).

Figure 3.8: AR52 tattooed pigskin before and after irradiation.

Pigskin samples were extracted with solvent, according to the procedure described in Paragraph 3.2.3: the extracts were analyzed by the micro-LC-MS/MS method developed; the generated data were analyzed using PCA. The dataset consisted of 102 samples (3 non-degraded pigskin, 3 degraded pigskin, 3 non-degraded tattooed pigskin, 3 degraded tattooed pigskin and the 90 samples of the previous dataset) and 2394 variables. A clear separation of the group of degraded pigskin occurs plotting PC1 versus PC2, but no difference between irradiated tattooed pigskin samples and irradiated non-tattooed pigskin was evident.

None of the degradation products found in the sulforhodamine B solutions could be identified in the degraded tattooed pigskin samples. In fact, PCA analysis evidenced only the degradation of the fatty acid chains present into the pigskin. In order to achieve information concerning AR52 degradation on tattooed pigskin, PCA-DA analysis was performed [43]. All the samples were correctly classified by the PCA-DA model: Figure 3.9 reports the loading and the score plots of the PC1 versus PC3.
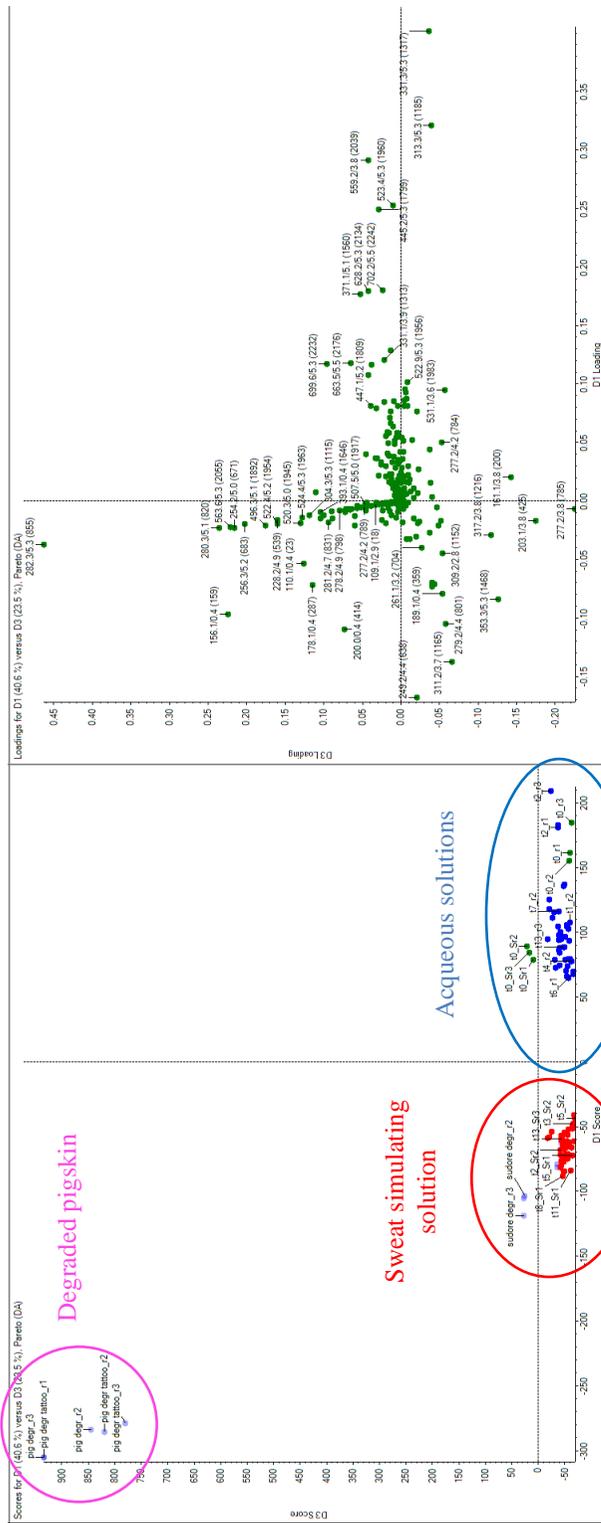
Figure 3.9: score (left) and loading (rigjt) plots of PC1 versus PC3 of the PCA-DA analysis model allowed the separation of threegroups of samples (acqueos solutions, sweat simulating solution and degraded pigskin samples).

109

Degraded pigskin samples are grouped at the left top (pink ellipse), while aqueous (blue ellipse) and sweat simulating solutions (red ellipse) lie at negative values on the y-axis of the score plot. The class of aqueous solutions is characterized by positive loading values on PC1, represented by their *m/z*, whereas the class of sweat-simulating solutions is at negative loading values on PC1. Despite this classification, it was still not possible to distinguish the irradiated tattooed pigskin samples from those irradiated but not tattooed. The main information explained by the PCA-DA model remains still related to the fatty acids degradation occurring in all the pigskin samples. To overcome the problem, hence to identify the degradation products formed on the irradiated tattooed pigskin, a *t*-test was adopted to compare the group of degraded non-tattooed pigskin samples with that of degraded tattooed pigskin samples. The *t*-test (at 95% confidence level) allows to understand if the compared data sets are statistically different for the tested characteristic. The *t*-test was applied to all the *m/z* intensities to determine whether the averages of the *m/z* intensities of each data set were significantly different from each other:

$$t_{95\%} = \frac{\sqrt{(\overline{G_1} - \overline{G_2})^2}}{s^2_{pooled} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

$\overline{G_1}$ = mean *m/z* intensity of the set of the irradiated tattooed pigskin samples;

$\overline{G_2}$ = mean *m/z* intensity of the set of the irradiated nontattooed pigskin samples;

n_1 = number of samples in group 1;

n_2 = number of samples in group 2;

s^2_{pooled} = pooled variance of the two sets, calculated as:

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where

$s_1=$ standard deviation of the *m/z* intensity of set 1;

$s_2=$ standard deviation of the *m/z* intensity of set 2.

As a result, a separated *t*-test was applied to each *m/z* value to discriminate the two sets of degraded pigskin samples. It was possible to highlight the *m/z* related to the dye degradation since the fatty acids degradation was common to both the tested data sets. In particular, a new signal at *m/z* 637.3051 (PI), labeled as P6, was evidenced only in degraded tattooed pigskin samples. Through the use of Peak-View[TM] and MarkerView, setting the same parameters adopted for degradation products and impurities identification, it was possible to attribute the chemical structure of P6 (Fig. 3.10).
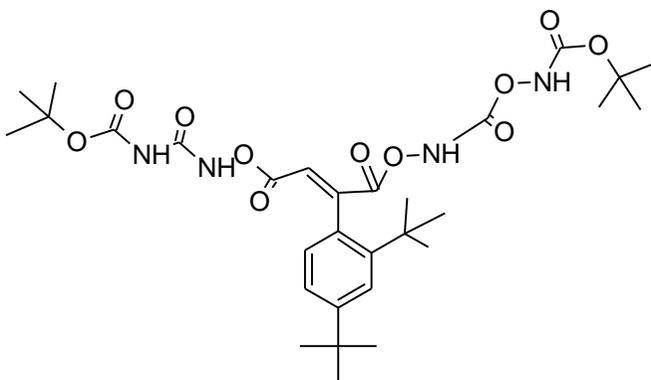

Figure 3.10: proposed structure of the ion at *m/z* 637.3051.

P6 is probably due to the reaction occurring between AR52 and pigskin, when subjected to solar irradiation. P6 is characterized by the opening of the xanthenic core and by the substitution of the two sulfonic groups by *tert*-butyl groups.

3.3.7 Evaluation of degradation products mutagenicity by the QSAR approach

Considering the hypothesized chemical structures of the degradation products, we evaluated their possible mutagenicity through a QSAR approach, commonly adopted by the US Environmental Protection Agency. The software program used to this aim was T.E.S.T. 4.2, able to evaluate the toxicity of a compound given its molecular structure [44-48]. QSAR model descriptors are hydrophobicity, electronic properties, steric hindrance, shape and size of the molecule. The Ames test uses *Salmonella typhimurium* characterized by mutations in genes that inhibit histidine synthesis attitude. If the strains are exposed to a mutagenic compound, their ability to synthesize histidine is restored: the bacterial colony can grow in a histidine free medium. T.E.S.T. evaluates compounds mutagenicity considering their capability to return to a "prototrophic" state, where the simulated cells can grow in a medium without histidine. The data set adopted for the QSAR model is constituted by 5743 compounds [49]. Concordance, sensitivity and specificity are the three parameters adopted to express model prediction accuracy. Concordance is the fraction of the correctly predicted compounds; sensitivity is the fraction of the experimentally active compounds that are correctly predicted as active, while specificity represents the fraction of compounds that are experimentally inactive and are predicted to be non-active. The threshold value for discriminating a mutagenic compound from one non-mutagenic is 0.5. The results were negative for all the tested degradation products and impurities present in the dye, as reported in Table 3.5.

Compound	Mutagenicity value	Concordance	Sensitivity	Specificity	Result
AR52	0.43	0.33	0.33	NA	neg
I1	0.28	0.50	0.50	NA	neg
I2	0.07	0.70	0.67	1.00	neg
P1	0.21	0.80	0.78	1.00	neg
P2	0.13	0.80	0.67	1.00	neg
P3	0.11	0.70	0.60	0.80	neg
P4	0.48	0.56	0.43	1.00	neg
P5	-0.03	0.80	0.75	0.83	neg
P6	0.03	0.90	0.75	1.00	neg

Table 3.5: virtual Ames mutagenicity evaluation of AR52 degradation products and impurities.

3.4 Conclusions

This study demonstrated that AR52 undergoes degradation when exposed to simulated sunlight both in aqueous and sweat-simulating solutions. The five degradation products in which AR52 turns into were identified by micro LC-MS/MS, and were common to both the tested aqueous media. AR52 degradation and discoloration rate was comparable with the permanence of nonpermanent tattoos applied on skin. None of the degradation products identified in aqueous and sweat simulating solutions were evidenced in samples of tattooed pigskin subjected to irradiation, but a signal at m/z 637.3051 demonstrated to be characteristic of the AR52 degradation on pigskin. A virtual mutagenicity test based on EPA QSAR model gave negative results for all the tested chemical structures, evaluating not only the degradation products, but also the two impurities present in the dye.

The performing of degradation studies on real samples, and not only on model solutions, demonstrated to be essential to understand the real situation in which

113

AR52 is applied on skin. Only in this way, in fact, it was possible to evaluate the possible interactions that may occur between the skin and the dye.

References

1. Forte G, Petrucci F, Cristaudo A, Bocca B (2009) Sci Total Environ 407: 5997-6002.

2. Braun K M, Perlmutter P, McDermott R J (2000) J Sch Health 70: 355-560.

3. Drews D R, Allison C K, Probst J R (2000) Psychol Rep 86: 475-481.

4. Marcoux D (2000) Dermatol Clin 18: 667-673.

5. Ceniceros S (1998) J Nerv Ment Dis 186: 503-504.

6. Greif J, Hewitt W (1998) Adv Nurse Pract 6: 26-31.

7. Armstrong M L, Masten Y, Martin R (2000) MCN Am J Matem Child Nurs 25: 258-261.

8. Armstrong M L, Murphy K P, Sallee A, Watson G (2000) Mil Med 165: 135-141.

9. Anderson R R (2001) Arch Dermatol 137: 210-212.

10. Ferguson J E, Andrew S M, Jones C J P (1997) Br J Dermatol 137: 405-410.

11. Timko A L, Miller C H, Johnson F B, Ross E (2001) Arch Dermatol 137: 143-147.

12. Friedman T, Westreich M, Mozes S N, Dorenbaum A, Herman O (2003) Plast Reconstr Surg 111: 2120-2122.

13. Moehrle M, Blaheta H J, Ruck P (2001) Dermatology 203: 342-344.

14. Mangas C, Fernandez-Figueras MT, Carrascosa J M, Soria X, Paradelo C, Ferrendiz C, Just M (2007) Dermatol Surg 33: 766-767.

15. Cui Y, Spann A P, Couch L H, Gopee N V, Evans F E, Churchwell M I, Williams L D, Doerge D R, Howard P C (2004) Photochem Photobiol 80: 175-184.

16. Vasold R, Naarmann N, Ulrich H, Fischer D, Könlg B, Landthaler M, Bäumler W (2004) Photochem Photobiol 80: 185-190.

17. Engel F, Santarelli F, Vasold R, Maisch T, Ulrich H, Prantl L, Köning B, Landthaler M, Bäumler W (2008) Contact Dermatitis 58: 228-233.

18. Engel F, Santarelli F, Vasold R, Ulrich H, Maisch T, Köning B, Landthaler M, Gopee N V, Howard P C, Bäumler W (2006) Anal Chem 78: 6440-6447.

19. Engel E, Spannberger A, Vasold R, Köning B, Landthaler M, Bäumler W (2007) J Dtsch Dermatol Ges 5: 583-589.

20. de Groot A C (2013) Contact Dermatitis 69: 1-25.

21. Rastogi S C, Johansen J D (2005) Contact Dermatitis 53: 207-210.

22. Onder M (2004) J Cosmet Dermatol 2: 126-130.

23. Madan V, Beck M H (2005) Contact Dermatitis 53: 241-242.

24. Hayakawa R, Fujimoto Y, Kaniwa M (1994) Am J Contact Dermat 5: 34-37.

25. Gosetti F, Chiuminatto U, Mazzucco E, Calabrese G, Gennaro M C, Marengo E (2012) Anal Chim Acta 746: 84-89.

26. Gosetti F, Chiuminatto U, Mazzucco E, Calabrese G, Gennaro M C, Marengo E (2013) Food Chem 136: 617-623.

27. Gosetti F, Gianotti V, Angioi S, Polati S, Marengo E, Gennaro M C (2004). J Chromatogr A 1054: 379-387.

28. Gianotti V, Angioi S, Gosetti F, Marengo E, Gennaro M C (2005) J Liq Chromatogr Relat Technol 28: 923-937.

29. Gosetti F, Gianotti V, Polati S, Gennaro M C (2005) J Chromatogr A 1090: 107-115.

30. Gosetti F, Frascarolo P, Mazzucco E, Gianotti V, Bottaro M, Gennaro M C (2008) J Chromatogr A 1202: 58-63.

31. Scientific Committee on Cosmetic Products (2004) Opinion of the Scientific Committee on Cosmetic Products and Non-Food Products Intended for Consumers concerning Acid Red 52. SCCNFP/0803/04.

32. Sosted H, Basketter D A, Estrada E, Johansen J D, Patlewicz G Y (2004) Contact Dermatitis 51: 241-254.

33. Gosetti F, Bolfi B, Marengo E (2015) Anal Bioanal Chem 407: 4649-4659.

34. Piunti F, Mucci N, Innocenti R, Fornaro G, Colonna G M, Bartolini G, Marchesi E (2004) Prev Oggi 2: 51-66.

35. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (1999) Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS) Umetrics; Scaling; 213-225.

36. Rajalakshmi S, Pitchaimuthu S, Kannan N, Velusamy P (2017) Appl Water Sci 7: 115-127.

37. Patrik W, Dietmar S (2007) Photochem Photobiol A 185: 19-25.

38. Keum Y S, Kim J H, Li Q X (2003) J Photosci 10: 219-223.

39. Natarjan T S, Thomas M, Natarjan K, Bajaj H C, Tayade R J (2011) Chem. Eng J 169: 126-134.

40. He Z, Yang S, Yu Y, Sun C (2009) J Environ Sci 21: 268-272.

41. Yu K, Yang S, He H, Sun C, Gu C, Ju Y (2009) J Phys Chem A 113: 10024-10032.

42. Ravera M, Musso D, Gosetti F, Cassino C, Gamalero E, Osella D (2010) Chemosphere 79: 144-148.

43. Gosetti F, Chiuminatto U, Mazzucco E, Mastroianni R, Marengo E (2015) Food Chem 167: 454-462.

44. Benigni R, Bossa C (2008) J Chem Inf Model 48: 971-980.

45. Modi S, Li J, Malcomber S, Moore C, Scott A, White A, Carmichael P (2012) J Comput Aided Mol Des 26: 1017-1033.

46. Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T, Singer T (2011) Chem Res Toxicol 24: 843-854.

47. Votano J R, Parham M, Hall L H, Kier L B, Oloff S, Tropsha A, Xie Q, Tong W (2004) Mutagenesis 19: 365-377.

48. Cassano A, Raitano G, Mombelli E, Fernàndez A, Cester J, Roncaglioni A, Benfenat E (2014) J Environ Sci Health C 32: 273-298.

49. US Environmental Protection Agency (2012) User's guide for T.E.S.T. (version 4.1) (Toxicity Estimation Software Tool). http://www.epa.gov/nrmrl/std/qsar/TEST-user-guide-v41.pdf.

Chapter 4

UHPLC-MS/MS target and non-target analyses of extra virgin olive oil samples stored in different packaging materials

3.2 Introduction

Many of the healthy virtues of the Mediterranean diet are associated to the consumption of Extra Virgin Olive Oil (EVOO), obtained from the first cold press of the fruits of the olive trees only throughout mechanical processes consisting of washing, decanting, centrifuging and filtrating [1]. EVOO healthy properties are mainly due to its minor components, constituted by polyphenols (PPs), considered as an important parameter to evaluate its quality [2-8]. From the chemical point of view, these compounds show a structure characterized by one or more hydroxyl or phenol groups, that are able to prevent oxidation reactions, protecting olive oil from rancidness, hence contributing to its stability [9-12]. In particular, olive oil stability is due for 30% to the PPs [13]. PPs act as antioxidants also in the living organisms, by inhibiting radical reactions that could lead to the formation of compounds able to alter the correct functioning of cells and tissues. The complex composition in PPs (phenolic acids such as caffeic acid, vanillic acid, syringic acid; lignans such as pinoresinol and 1-acetoxypinoresinol; hydroxytyrosol, tyrosol, and secoiridoids such as oleuropein aglycone, oleuropein, demethyloleuropein, ligstroside) also characterizes EVOO in terms of taste and flavor [14-18]. In particular, Regulation 432/2012 [19] certified the antioxidant and anti-inflammatory properties of PPs in olive oil. The Regulation reports that the claim "olive oil PPs contribute to the protection of blood lipids from oxidative stress" can

be used only for oils that contain more than 50 mg of hydroxytyrosol and its derivatives (oleuropein complex and tyrosol) for 20 g of oil.

Not only the olives variety and degree of maturation, tree location and environmental conditions, and the crop season [20], but also the different packaging material in which EVOO is stored [21], may highly influence the different amounts of the PPs in EVOO. In fact, packaging materials can alter their properties in terms of flavor and taste, but above all they can also release substances that could be toxic to humans [22-26]. In particular, Lavelli et *al*. [27] focused their investigation on the degradation of secoiridoids taking place in EVOO during storage. Their results showed that the EVOO samples characterized by high concentration of PPs maintained their beneficial properties during the period in which oil was stored in markets, even if some of the PPs degraded (i.e. oleuropein derivatives demonstrated to be less stable than the corresponding ligstroside derivatives). Conversely, the study conducted by Gómez-Alonso et *al*. [28] evidenced a PPs reduction ranging from 43% to 73% in oil samples stored at room temperature for 21 months: the highest values were obtained for the samples richer in PPs at the beginning of the study. Rizzo et *al.* [10] evaluated the conservation of EVOO in polyethylene terephthalate (PET) colored bottles, testing different stressing light conditions in order to simulate the market storage. One of the quality indexes most influenced by light conditions was indeed the PPs content. The researchers concluded that opaque (blue and white) plastic bottles preserves EVOO more than the colored ones, but that light conditions have to be carefully evaluated in order to extend the shelf life of the product. Riahi and Marzouk evaluated the quality of virgin olive oil packaged in dark glass bottles stored in dark conditions and in clear glass bottles subjected to daylight and non-stop artificial light [29]. Furthermore, Mendez et *al.* and Okogeri et *al.* [30, 31] studied how shelf life, effects of storage time and packaging materials (clear PET bottles, PET bottles covered with aluminum foil, glass, tin and Tetra-brik®) can influence virgin olive oil quality parameters. The quality of packaged EVOO is influenced by

the physico-chemical properties of the container material, its oxygen permeability, the content of oxygen in the headspace and the amount dissolved in the EVOO, temperature and light conditions and the original amount of antioxidants present in the EVOO.

Since PPs are an important index of EVOO quality, many analytical methods were developed to determine their content, considering different extraction methods, separation and quantification techniques [2, 17, 32-63]. Extraction is the step that influences the results more than other factors [63]: it must be easy, fast, not expensive, with a minimum impact on the environment and characterized by a good reproducibility and high recovery values. The most adopted extraction methodologies reported in literature are based on liquid-liquid extraction (LLE) and solid phase extraction (SPE). The liquid-liquid microextraction (LLME) proposed by Pizarro et *al.* [34] was further improved by Becerra et *al.* [47] adopting a smaller amount of sample, thus producing a smaller volume of residues.

Folin-Ciocalteu [49] method is one of the first techniques adopted to determine the overall PPs, but it lacks of specificity and sensitivity as it is a colorimetric assay based on the reaction of all oxydable phenolic hydroxyl group. Successively, other analytical techniques were developed to determine the PPs content. When GC is used [37, 38, 61] a derivatization step is required, while the drawback of CE methodologies is the low concentration sensitivity [50, 51, 60-62]. Conversely, the use of HPLC with UV [2, 33, 36, 39, 44, 61], fluorescence [36, 39] and MS detectors [1, 33, 39, 44, 47, 48, 52, 55, 58, 59], provides better results considering the analytical speed and cost reductions. Further improvements were achieved thanks to the UHPLC technique: the hyphenation of UHPLC with ESI MS provides methods characterized by high sensitivity and short analysis time [2, 39, 47, 52, 53, 63].

The presented study is part of a project committed by Assovetro[TM], whose aim was the monitoring of physical-chemical and nutritional properties of EVOO samples

stored in different packaging materials (dark glass, clear glass, bag-in-box, plastic and tin), considering both ageing and storage effects on the EVOO samples [64].

In particular, the part presented here was devoted to the investigation of the possible variations of the concentration of nine selected target compounds (Fig. 4.1) (tyrosol, hydroxytyrosol, quercetin, oleuropein, catechin, rutin, epigallocatechin gallate, epicatechin and pantothenic acid) as a function of the storage in the different containers along time. For this purpose a new UHPLC-MS/MS method was developed.
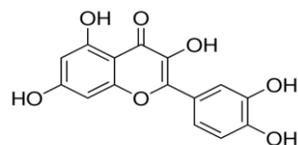
Furthermore, each sample was also investigated by non-target approach through micro liquid chromatography coupled with tandem high-resolution mass spectrometry, in order to identify possible markers typical of each of the tested packaging material.
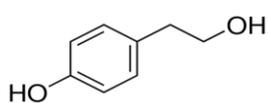
Tyrosol (TYR)
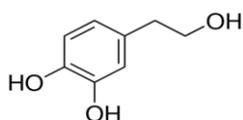
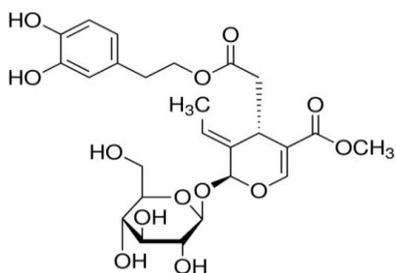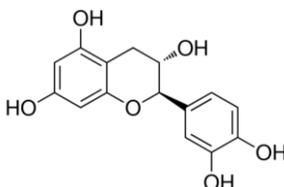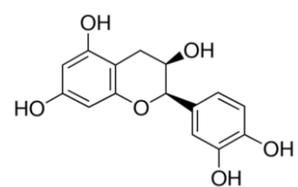**Hydroxytyrosol (HT)**

Quercetin (QUE)
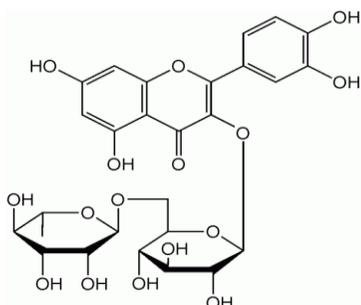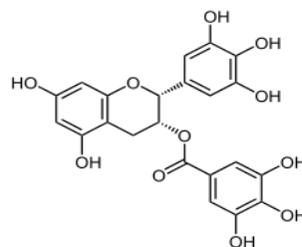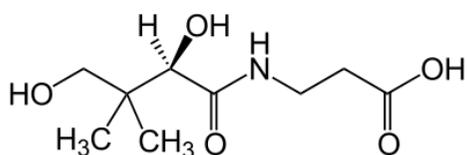
Oleuropein (OLE)

Catechin (CAT)

Epicatechin (EPI)

Rutin (RUT)

Epigallocatechin gallate (EGCG)

Pantothenic acid (B5)

Figure 4.1: chemical structures of the investigated analytes.

4.2 Experimental
4.2.1 Reagents

Reagent	Purity	Vendor
Methanol	LC-MS Ultra CHROMASOLV, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Ultrapure Water	For mass spectrometry	Fluka (Buchs, Switzerland)
Ethanol	Absolute, ≥ 99.8%	Sigma–Aldrich (Milwaukee, WI, USA)
Hexane	> 97 %	Sigma–Aldrich (Milwaukee, WI, USA)
Acetic acid	Eluent additive, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Formic acid	Eluent additive, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Hydroxytyrosol (HT)	> 98%	Sigma–Aldrich (Milwaukee, WI, USA)
Catechin (CAT)	> 96%	Sigma–Aldrich (Milwaukee, WI, USA)
Emodin (EMO)	> 97%	Sigma–Aldrich (Milwaukee, WI, USA)
Epicatechin (EPI)	> 98%	Sigma–Aldrich (Milwaukee, WI, USA)
Epigallocatechin gallate (EGCG)	> 95%	Sigma–Aldrich (Milwaukee, WI, USA)
Oleuropein (OLE)	> 98%	Sigma–Aldrich (Milwaukee, WI, USA)
Pantothenic acid (B5)	> 98%	Sigma–Aldrich (Milwaukee, WI, USA)
Quercetin (QUE)	> 98%	Sigma–Aldrich (Milwaukee, WI, USA)
Rutin trihydrate (RUT)	> 95%	Sigma–Aldrich (Milwaukee, WI, USA)
Tyrosol (TYR)	> 99.5%	Sigma–Aldrich (Milwaukee, WI, USA)

Table 4.1: list of the used reagents.

4.2.2 Target analysis apparatus

The UHPLC-MS/MS target analyses were performed on a Nexera LC system (Shimadzu, Kyoto, Japan) equipped with a DGU-20A3R Degasser, two LC-30AD Pumps, a SIL-30AC Autosampler, a CTO-20AC column compartment and a CMB-20A Lite system controller. The system was coupled with a 3200 QTrap[TM] LC-MS/MS system (Sciex, Concord, Canada) by a Turbo V[TM] interface equipped with an ESI probe. The 3200 QTrap[TM] data were processed by Analyst 1.5.2 software (Sciex, Concord, Toronto, Canada).

4.2.3 Non-target analysis apparatus

Non-target analyses were performed using a micro LC system (Eksigent Technologies, Dublin, CA, USA) constituted by an Eksigent Technologies micro LC200 pump with a flow module of 5-50 µL and a CTC programmable array logic autosampler with a Peltier unit (1-45 °C). The system was coupled with a TripleTOF[TM] system (Sciex, Concord, Canada) equipped with a DuoSpray[TM] ion source and calibrant delivery system. PeakView[TM] 2.1 (Sciex, Concord, Canada) was adopted to investigate accurate mass spectral data, while MasterView[TM] 1.0 (Sciex, Concord, Canada) was used to compare the intensities of the mass spectral data along the different samples.

4.2.4 Stock solutions and sample preparation

Each standard stock solution was prepared in methanol at the concentration of 100.00 mg L^{-1} and preserved at -20 °C in dark glass vials. The prepared solutions were stable for ten months.

Italian EVOO samples were provided by the same producer that bottled the EVOO, coming from the same pressing, in the five different packaging materials considered (dark glass, indicated as DG; clear glass, indicated as CG; tin, indicated as TN; plastic, indicated as PL, bag-in-box, indicated as BB). In order to

investigate only the variations due to the packaging material, the samples were purchased at the same time, and stored under the same temperature and light conditions until the end of the study. Packaged EVOO was stored at room temperature 22 ± 2 °C for ten months. The PPs content was determined on a monthly basis, both on containers opened, analyzed and resealed (*ageing*), and on sealed containers, opened and immediately analyzed (*conservation*).The a*geing* study considered samples analyzed at 0, 6, 13, 20, 27, 55, 83, 111 and 139 days from the opening of the same EVOO contained in the five containers, whereas *storage* study considered samples opened and immediately analyzed at 0, 6, 13, 20, 27, 55, 83, 111, 139, 167, 196, 230, 251, 286 and 314 days after bottling. 115 EVOO samples were considered, evaluating 15 sampling times for the conservation study and 9 sampling times for the ageing study, for each of the five different packaging materials.

Two subsequent liquid-liquid extractions (LLE) spaced out by two centrifugations (8000 rpm at 4 °C for 5 min) were performed on 2.00 g of EVOO. The first used a mixture 70/30 v/v ethanol/water (2.000 mL), while the second employed pure hexane (2.000 mL), in order to remove the fatty moiety. The aqueous extract was collected and successively diluted 1:3 v/v in water with 0.01% acetic acid before the injection into the UHPLC-MS/MS system.

4.2.5 Chromatographic and mass spectrometric conditions for target analysis

An Acquity UPLC HSST3 (2.1 mm x 100 mm, 1.8 μm, Waters, Italy) was selected as the stationary phase. The mobile phase was a mixture of ultrapure water with the addition of 0.01% acetic acid (A) and methanol with the addition of 0.01% acetic acid (B), in the following gradient (Tab. 4.2):

Time (min)	Flow rate (μL min^{-1})	% A	% B
0.0	0.4	80.0	20.0
0.5	0.4	80.0	20.0
3.0	0.5	35.0	65.0
3.1	0.5	0.0	100.0
7.0	0.5	0.0	100.0
7.1	0.4	80.0	20.0
10.0	0.4	80.0	20.0

Table 4.2: target analysis elution gradient.

The injection volume was 5.0 μL and the oven temperature was set at 40 °C.

The Turbo V^{TM} ion source worked in NI mode, setting the following conditions: curtain gas (N_2) at 30 psig, nebulizer gas GS1 and GS2 at 50 and 40 psig respectively, desolvation temperature (TEM) at 550 °C, collision activated dissociation gas (CAD) at 6 units of the arbitrary scale of the instrument and ion spray voltage (IS) at -4500 V. Mass spectrometer worked in scheduled multiple reaction monitoring (sMRM) considering the transitions of each analyte at a prefixed retention time.

Unit mass resolution was established and maintained in each mass-resolving quadrupole by keeping a full width at half maximum (FWHM) of about 0.7 u.

4.2.6 Chromatographic and mass-spectrometric conditions for non-target analysis

The stationary phase was a Halo C18 (0.5 mm x 100 mm, 2.7 μm; Eksigent Technologies, Dublin, CA, USA), while the mobile phase was constituted by a mixture of ultrapure water with the addition of 0.1% (v/v) formic acid (A) and

methanol with the addition of 0.1% (v/v) formic acid (B), eluting at flow rate 20 μL min^{-1}, at the following gradient conditions (Tab. 4.3):

Time (min)	% A	% B
0.0	90.0	10.0
0.5	90.0	10.0
4.0	40.0	60.0
5.5	0.0	100.0
12.0	0.0	100.0
12.1	90.0	10.0
15.0	90.0	10.0

Table 4.3: non-target analysis elution gradient.

The injection volume was 300 nL and the oven temperature was set at 45 °C.

The DuoSpray^{TM} ion source worked both in positive ion (PI) mode and negative ion (NI) mode. The instrumental parameters were set as follows: curtain gas (N_2) at 35 psig, nebulizer gas GS1 and GS2 25 and 35 psig respectively, desolvation temperature at 350 °C, ion spray floating voltage at ±5500 V depending on the polarity adopted. The instrument worked in Independent Scan Mode through Sequential Windowed of All Theoretical MS (SWATH-MS/MS^{ALL}) acquisition. A single TOF scan in a mass range of 50-800 Da (accumulation time 250 ms), was followed by MS/MS scans (accumulation time 20 ms) of 70 isolation windows of 11 Da in the range of 100-800 Da in high sensitivity mode. Consecutive isolation windows were overlapped of one mass unit. SWATH acquisition provided a comprehensive analysis of all the EVOO samples, through the full scan high resolution MS/MS acquisition on all the detectable compounds eluting from the chromatographic column. The Data Independent Scan Mode acquisition gives the

complete fingerprint of the analyzed samples, allowing the detection and identification even of molecules present in small amounts.

4.2.7 Statistical data analysis

PCA calculation was performed by Statistica v.7.1 (Statsoft Inc, Tulsa, OK, USA), while PLS-DA was performed by Parvus (developed by Prof. Michele Forina, Pharmacy Department, Genoa) and The Unscrumbler X 10.3 (Camo, Oslo, Norway) and some routines self-developed, in Matlab environment (Matlab R2014b, The Mathworks, Natik, MA, USA).

4.3 Results and discussion

4.3.1 Mass spectrometry characterization of the analytes

Each analyte was subjected to a MS/MS characterization study in order to identify its fragmentation pattern when the collision energy was increased. The experiments were performed by direct infusion of 1.0 mg L^{-1} standard solution of each analyte trough a T valve to the syringe pump (syringe flow rate: 10.0 µL min^{-1}, chromatographic pump flow rate 100 µL min^{-1}). All the compounds were characterized by many transitions: the two most intense ("quantifier" and "qualifier", respectively) were selected to respectively identify and confirm the presence of a specific analyte. Table 4.4 reports the optimized instrumental parameters for each compound.

Analytes	Q1 (m/z)	Q3 (m/z)	tR (min)	DP (V)	EP (V)	CEP (V)	CE (V)	CXP (V)
HT	153	123/93	1,5	-34	-7	-17.317	-20/-31	-1.08/-1.35
B5	218	88/146	1,3	-36	-4	-19.722	-18/-22	-1.21/-0.69
CAT	289	109/123	1,9	-48	-2	-22.349	-32/-37	-0.58/-1.01
EGCG	457	169/125	2,5	-26	-8	-28.569	-23/-53	-1.14/-0.74
Emo (IS)	269	239/211	4,3	-45	-7	-21.613	-42/-54	-1.82/-1.64
EPI	289	109/123	2,8	-48	-2	-22.349	-32/-37	-0.58/-1.01
OLE	539	275/307	3,9	-40	-5	-31.606	-28/-29	-1.89/-2.12
QUE	301	151/179	4,0	-59	-4	-22.793	-30/-23	-0.78/-1.16
RUT	609	300/271	3,8	-74	-7	-34.2	-46/-80	-1.90/-1.90
TYR	137	106/119	2,3	-27	-10	-16.725	-23/-24	0.00/0.00

Table 4.4: sMRM NI transitions (Q1 and Q3 masses), retention time (t_R), and mass spectrometry parameters: DP (Declustering potential), EP (Entrance Potential), CEP (Collision cell Entrance Potential), CE (Collision Energy), CXP (Collision cell eXit Potential). The most two sensitive transitions for each species were monitored.

4.3.2 Development and optimization of the LC-MS/MS target analytical method

Different stationary and mobile phases were tested: the obtained results were compared evaluating chromatographic resolution, peak symmetry and run time.

The first stationary phase tested was a Kinetex C18 column (2.1 mm x 150 mm, 1.7 μm, Phenomenex, Italy), but it did not allow the complete separation of all the analytes since some of their peaks were broad, while others showed a low intensity. Since the phenyl rings present in the phenyl columns improve selectivity towards polar compounds like the PPs, an Accucore phenyl-hexyl column (2.1 mm x 100 mm, 2.6 μm, ThermoFisher Scientific, Italy) was tested using the same mobile phase adopted for Kinetex C18. Also this stationary phase did not improve the peaks shapes: most of the peaks were still broad and some analytes were co-eluting (i.e. hydroxytyrosol and pantothenic acid, oleuropein and quercetin).

In order to achieve a satisfactory chromatographic separation of all the analytes, a Kinetex PFP column (2.1 mm x 150 mm, 1.7 μm, Phenomenex, Italy) was then tested. The choice was driven by the particular bonded phase that characterizes the PFP columns: the fluorine atoms on the phenyl ring make the column highly suitable for polar compounds containing hydroxyl and carboxyl groups. Unfortunately, also this stationary phase did not provide the desired chromatographic separation, since it presented the same problems of the previous ones.

Successful results were achieved using a HSST3 column (2.1 mm x 100 mm, 1.8 μm, Waters, Italy): its C18 trifunctionalized alkyl phase provided better retention, hence better separation, of the polar compounds and improved the compatibility with aqueous mobile phases. A valley-to-valley separation of all the analytes was achieved, and peak shapes improvements were obtained, thus HSST3 was selected as the stationary phase.

To improve the analytes separation maximizing their response during the ionization process and providing shorter analysis time, different mixtures of water, methanol and acetonitrile were tested, also evaluating possible improvements due to the addition of acetic or formic acid (both tested in the range between 0.01% and 0.1% v/v) or organic salts (ammonium acetate and ammonium formate were tested in the concentration range between 1.0 and 20.0 mM).

The mixture of water and methanol both with the addition of formic acid in any of the tested percentages did not allow the detection of tyrosol: too low mobile phase pH values, in fact, hinder its ionization. Since formic acid was not suitable, acetic acid was tested at different percentages. Percentages greater than 0.01% v/v led to a significant reduction of the ionization of tyrosol.

Acetonitrile with the addition of 0.01% of acetic acid was also tested, but it provided worst results than methanol since it did not promote tyrosol ionization process as methanol did. The addition to the mobile phases of organic salts at different concentrations was then tried, but it did not lead to significant improvements. For these reasons, the selected mobile phase was a mixture of ultrapure water and methanol, both with the addition of 0.01% acetic acid.

Figure 4.2 reports the chromatographic separation of all the analytes at the selected experimental conditions: a valley-to-valley separation was obtained for all the chromatographic peaks and emodin (employed as internal standard) was the last eluted compound (t_R=4.3 min). The remaining run time was used by the system to clean the column with 100% of methanol for 4 minutes at a flow rate of 0.500 μL min^{-1}, in order to remove all the possible lipophilic species present in EVOO, hence preserving the column life time.
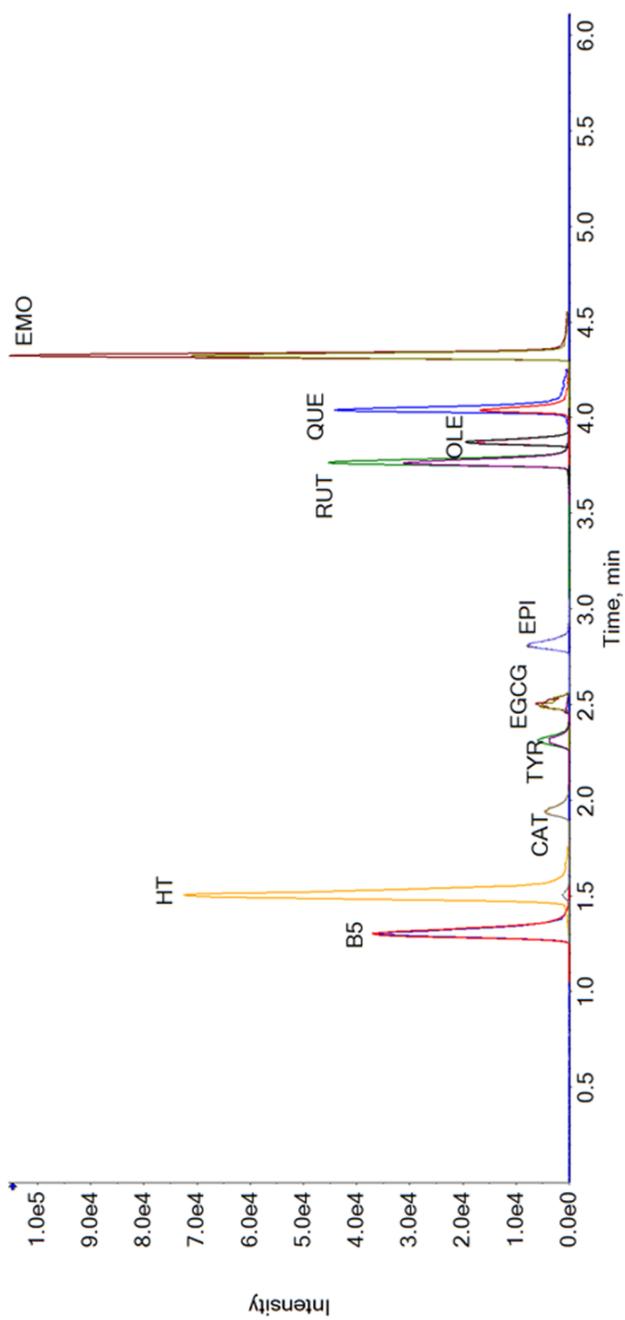
Figure 4.2: chromatographic separation of the analytes (100.0 µg L^{-1} each) in the optimal experimental conditions: quantifier and qualifier transitions are shown for each peak.

4.3.3 Development of the LC-MS/MS non-target analytical method

Chromatographic and mass spectrometric conditions were chosen in order to be as much unselective as possible, thus to explore all the possible classes of compounds present in the EVOO samples: both PI and NI modes were explored, since there was no *a priori* knowledge about the compounds present. SWATH[TM] protocol is a Data Independent Scan mode: it consists in the fragmentation of all the detectable ions with steps of 10 *m/z* isolation windows, (independently from their intensity). All the *m/z* range is scanned through subsequent isolation windows, exhaustively mapping each sample.

4.3.4 Validation of the analytical method

First a calibration plot was built for each analyte: on the *y*-axis is the peak area of the "quantifier" transition and on the *x*-axis is the concentration of the standard. Six concentration levels (each replicated three times) were analysed in the range between the LOQ value and 1000.0 μg L[-1]. The standard solutions were injected in randomized order to overcome possible memory effects. A 1/x weighted linear regression fit was adopted for all the target compounds. The calibration equations were all characterized by a good linearity, since all the obtained correlation coefficients (R^2) were always greater than 0.9938. To check the significance of the intercept of each calibration equation, a *t*-test (95% confidence level) was performed. The ratios between the calculated and the theoretical concentrations were evaluated using spiked samples at concentrations of 10.0, 50.0, 100.0, 250.0, 500.0, 1000.0 μg L[-1] of the analyte mixture, all prepared and analyzed in sets of five replicates. For the six concentration levels the obtained ratios were all in the range between 89% and 115%.

Limit of Detection (LOD) was calculated as the concentration of the analyte that gives a signal (peak area) equal to the average background (S_{blank}) plus three times the SD s_{blank} of the blank:

$$LOD = S_{blank} + 3s_{blank}$$

The Limit of Quantification (LOQ) was calculated as:

$$LOQ = S_{blank} + 10s_{blank}$$

Intra- and inter-day precisions were calculated both on retention time and concentration, analyzing a standard mixture of all the analytes (100.0 µg L^{-1} each) every day (five replicates) for a week. The obtained results showed intra- and inter-day RSD% calculated on retention time lower than 0.65% and 2.14%, respectively, whereas intra- and inter- day RSD% calculated on concentration were lower than 2.09 and 5.71%, respectively. An internal standard (IS) was adopted to check possible quantitative response changes: a 500 µg L^{-1} emodin solution was added to each solution prior the injection into the LC-MS/MS system. All the obtained results laid within ± 3σ control limits of the calibration plots. Possible memory effect were evaluated injecting blank samples after the injection of the highest concentration level of the standard mixture: the result was negative.

The recovery R was evaluated for each analyte in each sample: analyte mixture at three different concentration levels were added to the oil samples. In order to obtain concentrations laying in the linearity range, the solutions (50.0, 250.0 and 500.0 µg L^{-1}, each replicated three times) were prepared considering the 1:3 dilution v/v in ultrapure water with the addition of 0.01% acetic acid v/v before the injection. The recovery values were calculated as:

$$R = \frac{C_{obs}}{C_{ref}}$$

C_{obs} = difference between the concentration determined for the spiked sample and the native concentration in the sample;

C_{ref} = spiked concentration

A 95% confidence level *t*-test was performed to evaluate possible significant differences occurring between the R values obtained for the three concentration levels. The results demonstrated that no effect due to concentration was affecting recovery values. For this reason, recovery was expressed as an average percentage $\bar{R}(\%)$ for each analyte and calculated for each sample matrix. All the $\bar{R}(\%)$ ranged from 73.7% to 100.0%. To assess whether the quantitative obtained data had to be corrected for the $\bar{R}$ values, again a *t*-test at 95% confidence level was applied to determine if the $\bar{R}(\%)$ were statistically different from 100% [65]. The results indicated that no correction was necessary for TYR and OLE, since their recoveries were quantitative.

The Matrix Effect (*ME*) was evaluated both in terms of extent and presence. Different dilution ratios were tested (2, 3, 5, 10, 20) to achieve the best compromise between method sensitivity and matrix effect relevance. The best result was provided by the 1:3 dilution ratio.

In order to check the presence of the matrix effect, two sets of plots were built: external calibration plot and standard addition plot, built for all the real samples, spiking the samples with standard solution mixture at 50.0, 250.0 and 550.0 µg L^{-1}. A *t*-test at 95% confidence level compared the two slopes; the results were expressed as *ME* percentage (*ME%*) according to:

$$ME(\%) = \frac{slope_{add}}{slope_{ext}} \cdot 100 - 100$$

When the result of the *t*-test (*ME%)* is equal to zero, no *ME* affects the results, while a negative value indicates signal suppression and a positive one suggests signal enhancement.

The results evidenced that none of the investigated analytes was affected by *ME*.

Linearity range, R^2 of the calibration plots, LODs, LOQs, intra- and inter- day RSD% values calculated on retention time and on concentration and recovery are reported in Table 4.5.

Name	Range	R^2	LOD (µg/L)	LOQ (µg/L)	RSD % (t_R) intra day (n=5)	RSD % (t_R) inter day (n=35)	RSD % (conc) intra day (n=5)	RSD % (conc) intra day (n=35)	Recovery (%)
HT	1.1-1000.0	0.9986	0.3	1.1	0.49	1.17	0.79	1.02	88.4±7.2
B5	1.1-1000.0	0.9998	0.3	1.1	0.65	2.14	1.14	2.20	87.7±6.9
CAT	2.9-1000.0	0.9996	0.9	2.9	0.42	0.77	1.57	2.07	73.7±6.7
EGCG	28.3-1000.0	0.9962	8.5	28.3	0.30	0.56	2.09	4.34	100.0±2.3
EPI	2.0-1000.0	0.9980	0.6	2.0	0.20	0.42	1.53	3.15	79.1±7.3
OLE	1.6-1000.0	0.9982	0.5	1.6	0.10	0.12	1.17	1.90	99.8±8.5
QUE	1.1-1000.0	0.9938	0.3	1.1	0.01	0.10	1.50	5.71	87.9±4.8
RUT	0.8-1000.0	0.9984	0.3	0.8	0.14	0.39	0.95	4.28	87.6±4.0
TYR	1.4-1000.0	0.9998	0.4	1.4	0.29	0.42	1.14	1.56	97.4±9.1

Table 4.5: Linearity range, regression coefficient (R^2), LOD, LOQ, intra- and inter-day RSD (%) of retention time and of concentration and average recovery for each analyte.

4.3.5 UHPLC-MS/MS target analysis of extra virgin olive oil samples

The developed UHPLC-MS/MS method was applied to the 115 EVOO samples. Figure 4.3 shows the chromatogram acquired for an EVOO sample stored in dark glass at the beginning of the study, spiked with a mixture of all the analytes (each at 50.0 µg L^{-1}). Table 4.6 reports the concentration of hydroxytyrosol, quercetin and tyrosol (µg Kg^{-1}): other investigated analytes are not reported because they were not detected.
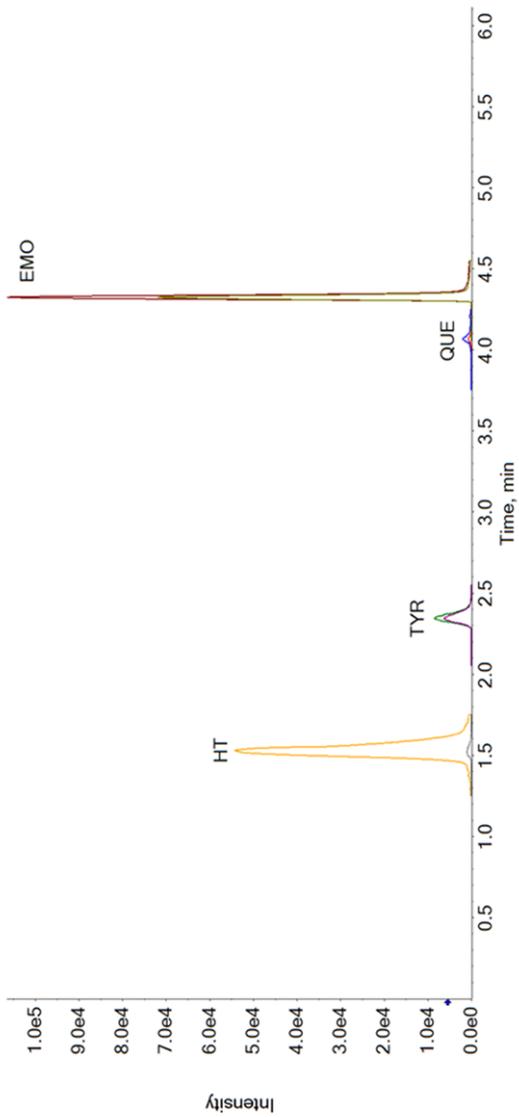
Figure 4.3: chromatogram of a EVOO sample stored in dark glass at the beginnings of the study, in which hydroxytyrosol, tyrosol and quercetin were quantifieed.

hydroxytyrosol	quercetin	tyrosol
407.2±4.8	15.3±7.3	474.5±27.7

Table 4.6: concentration of hydroxytyrosol, quercetin and tyrosol in EVOO sample in dark glass at the beginning of the study.

Tyrosol and hydroxytyrosol are the most abundant analytes present in the samples. This is an expected result since, to be considered extra virgin and to provide the typical health-promoting properties, olive oil must contain at least a concentration of 5 mg Kg^{-1} of hydroxytyrosol and its derivatives (oleuropein complex and tyrosol) for 20 g of olive oil [66]. No significant trends were observed for all the analytes considering both ageing and conservation studies, except for quercetin (Fig. 4.4 and Fig. 4.5). Its concentration, in fact, decreases along time for all the packaging materials, but in a slower rate in dark glass (indicated by the blue line), suggesting a better protection of EVOO from external effects rather than in the other packaging materials. In particular, ageing study on quercetin stored in DG showed a fast decrease of its concentration up to day 20, then its level remained constant for the remaining days, at a level similar to the one of the other packaging materials (Fig. 4.4). Considering storage study, quercetin concentration slowly decreased in DG until day 111, then such trend became more evident up to day 196 and finally remained constant up to the end of the study. A similar, but less evident, trend was evidenced in TN, while quercetin content in BB, PL and CG remained constant for all the study at its already initial concentration (Fig. 4.5).
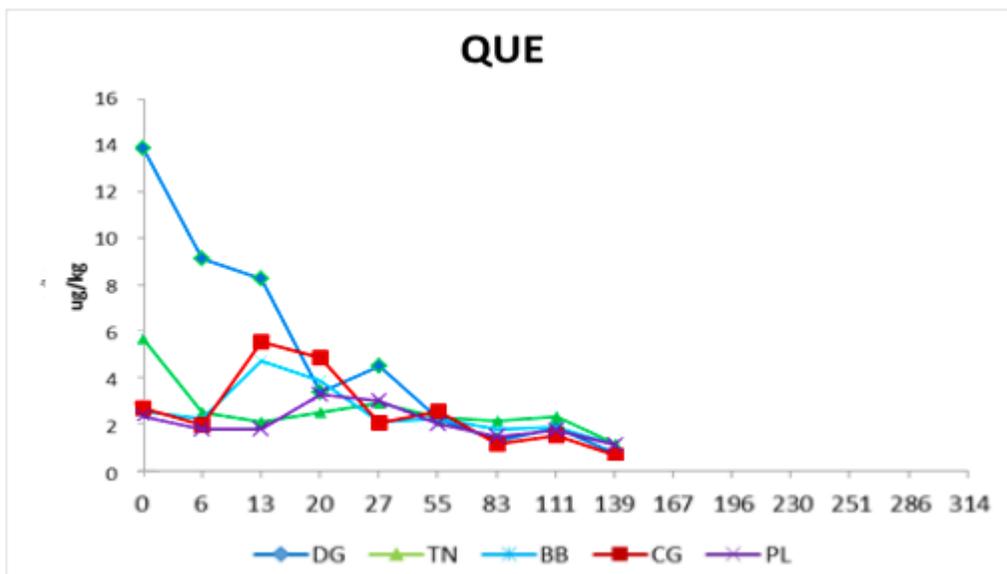
Figure 4.4: Quercetin concentration decreases in dark glass (DG), tin (TN), bag-in-box (BB), clear glass (CG) and plastic (PL) opened sealed containers and monthly analyzed (ageing).
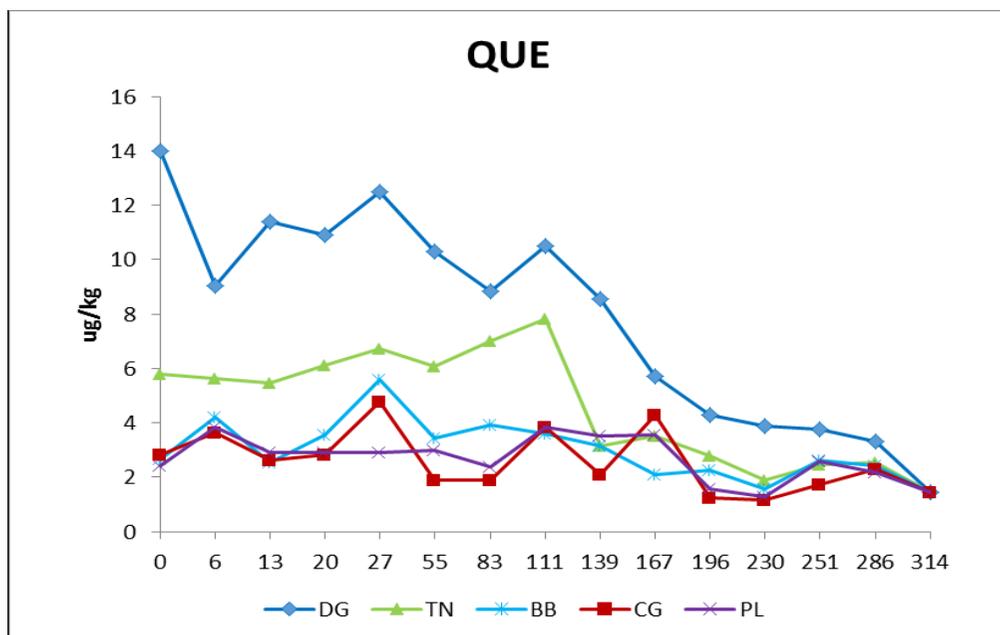

Figure 4.5: Quercetin concentration decreases in dark glass (DG), tin (TN), bag-in-box (BB), clear glass (CG) and plastic (PL) sealed containers monthly opened and immediately analyzed (conservation).

Tyrosol and hydroxytirosol trends along time are not reported here since it was not possible to observe particular trends ascribable to the different packaging material used.

In all the analyzed samples, pantothenic acid, epigallocatechin gallate, catechin, epicatechin, oleuropein and rutin were not detected.

4.3.6 Strategies for the identification of unknown species

Each EVOO sample was investigated also by a non-target approach, using the unselective LC-MS conditions described in Paragraph 4.3.3.

A comparative screening between EVOO samples and control (aqueous mobile phase) was performed in order to underline signals ascribable only to EVOO compounds. The workflow for the selection of non-target peaks was carried out in a semi-automated mode by the software PeakView^{TM}, after the setting of the following parameters:

- Do not calculate details for XIC with intensity < 1000 counts;
- Signal/noise ratio < 10;
- XIC width (Da) = 0.02;
- Retention time width (min) = 2;
- Threshold (cps) = 100;
- Threshold (ratio of control) = 10;
- Minimum retention time = 0.3 min;
- Maximum retention time = 13 min.

Furthermore, when all these parameters were overcome, the software was able to compare the intensity ratio of the two compared chromatographic runs, indicating as significant signals only those that were at least ten times more intense than the control sample. The low ratio of control value set was chosen in order to find out

all the possible compounds present in EVOO samples, even at very low concentration. Automatically, all the significant signals were included in a list and extracted from the TOF scan. The acquired MS and MS/MS spectra were used by MasterView^{TM} (a PeakView^{TM} application tool) to automatically calculate a possible molecular formula associated to each of the significant signals and search for any correspondence with the online database. In fact, the software is supported by the online ChemSpider database, useful to suggest other potential molecular and structural formulas for the significant *m/z*. In particular, it provides many structural suggestions for the same molecular formula. MasterView^{TM} compares the experimental MS/MS spectrum for the extracted *m/z* and that simulated *in silico* for the same compound by the software, based on the knowledge of breaking and rearranging bonds. The result is a matching percentage that indicates how the proposed structure is pertinent to the acquired spectra.

The described workflow allowed the attribution of the signal at *m/z* 303.1064 NI, present in all the investigated samples, to oleocanthal, a typical phenolic compound present in EVOO, confirmed by a matching percentage greater than 75%.

Since the peak list generated by PeakVieW^{TM} for each EVOO sample consisted in hundreds of significant signals, the described investigation procedure would have been inefficient and time consuming. In order to reduce the number of ions, significantly different from experimental error, statistical data processing algorithms approach were applied. Principal Component Analysis (PCA) was adopted to reduce the data complexity and highlight the *m/z* signals related to the differences occurring between the different materials in both storage and ageing studies. Before performing PCA, all the chromatographic runs were aligned and the data contained therein were autoscaled to ensure that all variables would have contributed with the same amount of information. The dataset associated to the conservation study (sealed containers monthly opened and immediately analyzed) consisted of 75 samples (15 sampling times for each of the 5 packaging material tested) and 17232 variables (13837 *m/z* signals deriving from positive ionization

and 3395 from the negative one). Figure 4.6A and 4.6B report the score plots of the first two PCs considering all the samples related to storage study grouped for kind of packaging material independently from sampling time (Fig. 4.6A) and samples grouped for sampling time independently from type of packaging material (Fig. 4.6B). From Figure 4.6A and 4.6B it is possible to see how sampling time is the preponderant information.
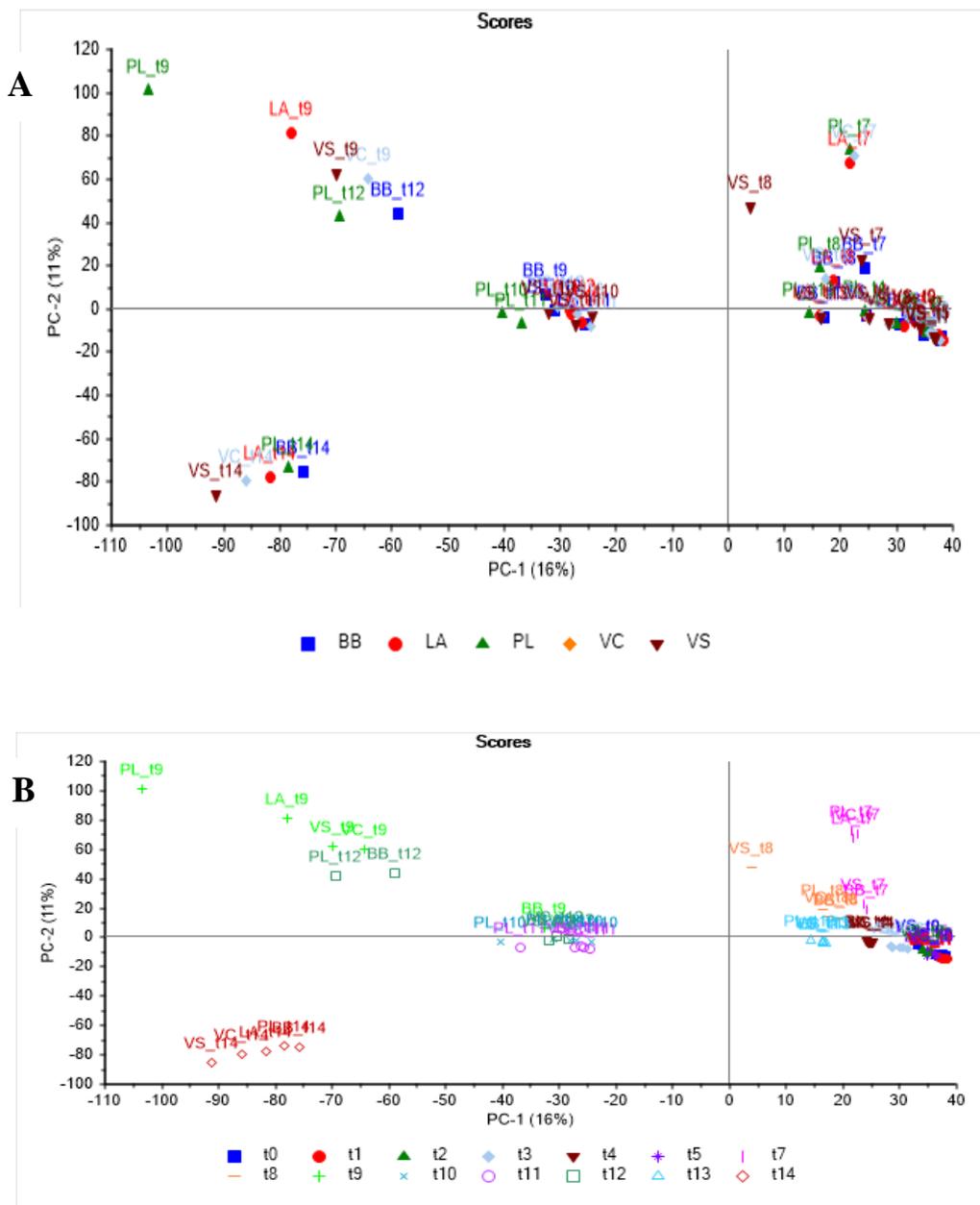
Figure 4.6: A) score plot of the first two PCs related to EVOO samples stored in the five packaging containers, monthly opened and immediately analyzed independently from the sampling time; B) score plot of the first two PCs related to the EVOO samples stored in the different packaging materials grouped for sampling time, independently from the type of packaging material.

The two score plots of the two first PCs did not show a clear separation of the samples stored in the five different packaging materials because the predominant information was related to the conservation time.

To obtain information related only to the packaging material, a supervised technique was adopted: Partial Least Squares-Discriminant Analysis (PLS-DA, described in Paragraph 1.5.2.2).

PLS-DA was applied selecting the variables through a backward selection algorithm: the algorithm eliminates one variable at each iteration, therefore the model consists of the variables that allow the maximum predictivity ability of the model, evaluated through a leave-one-out cross-validation (LOO). The results of the PLS-DA model built for the comparison of the EVOO samples stored in the different sealed containers monthly opened and immediately analyzed (*storage study*) are reported in Table 4.7.

Comparison	R^2	R^2_{CV} (LOO)	N° LV	N° m/z (NI) signals	N° m/z (PI) signals in the model	N° m/z (NI) signals	N° m/z (NI) signals in the model
BB vs CG	0.9180	0.7614	2	4416	140	1145	62
BB vs DG	0.8733	0.5877	2	4456	332	1166	106
TN vs CG	0.8598	0.7624	2	4567	26	1207	20
TN vs DG	0.9134	0.8098	2	4526	20	1204	20
PL vs CG	0.9917	0.8870	2	4774	76	1318	30
PL vs DG	0.9390	0.8050	2	4739	97	1348	48
CG vs DG	0.9145	0.7126	2	3935	151	967	55

Table 4.7: results of the PLS-DA classification method adopted to compare EVOO samples stored in clear glass (CG) and dark glass (DG) with those stored in bag-in-box (BB), tin (LA) and plastic (PL).

The information present in the samples was successfully described by the method since all the correlation coefficients in fitting (R^2) are greater than 0.85 and the correlation coefficients cross validated (R^2_{CV}), describing the predictivity of the model, are all greater than 0.76; the only exception is represented by the model that describes BB vs DG, whose R^2_{CV} is 0.5877. It has to be underlined that the number of significant signals (columns 5 and 7) was highly reduced (columns 6 and 8). Such evidence is coherent with the results obtained by the PCA that showed how the information related to ageing was more significant than that related to the type of material.

As an example, hereafter the comparison between clear glass (CG) and plastic (PL) will be discussed. Figure 4.7 shows the predicted and observed class belongings, evidencing how the EVOO samples stored in CG and in PL are correctly classified in the two different classes. The blue objects represent the fitting results, while the red ones those in cross-validation. The plot shows satisfactory results as the model is characterized by $R^2 = 0.9917$ and R^2_{CV} (LOO)= 0.8870.
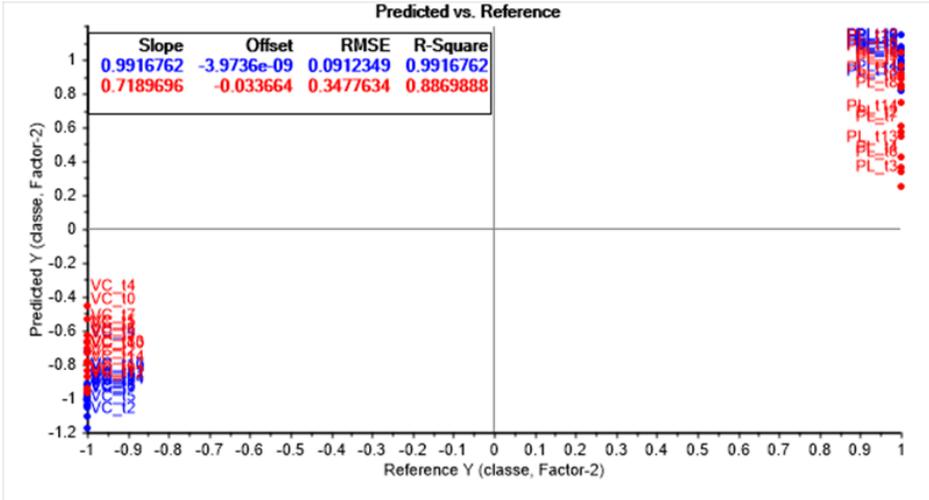

Figure 4.7: predicted (in red) and observed (in blue) objects for the samples stored in CG and PL.

The score plot reported in Figure 4.8 shows the objects in the area described by the first two latent variables, constituting together the 99% of the information related to the type of packaging material (Y variable) and the 16% of the acquired signals (X matrix).

The samples stored in CG (red ellipse) are clearly separated from those stored in PL (blue ellipse).
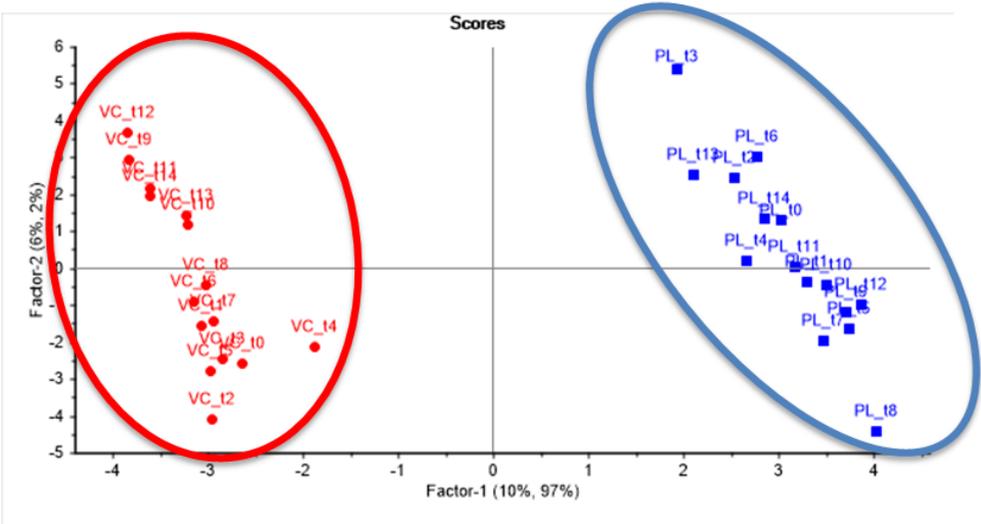

Figure 4.8: score plot of the first two latent variables that allow the separation of EVOO samples stored in CG (*red ellipse*) and those stored in PL (*blue ellipse*).

The loading plot reported in Figure 4.9 allows the identification of the variables that differentiate the samples stored in CG and PL. The variables circled in blue show greater values for samples stored in PL then in those stored in CG, while the variables circled in red show an opposite trend.

Figure 4.9: loading plot of the first two latent variables that allows the separation of the EVOO samples stored in CG from those stored in PL.

The dataset related to the ageing study (opened sealed containers and monthly analyzed) consisted of 45 samples (9 sampling times for each of the 5 packaging material tested) and 8582 variables (6267 *m/z* signals deriving from positive ionization and 2315 from the negative one). Even for the ageing study, PCA did not allow the clear separation of the five packaging materials since again the most relevant information was related to the sampling time. PLS-DA were again performed on the same data. Table 4.8 reports the results of this analysis.

Comparison	R^2	R^2_{CV} (LOO)	N° LV	N° m/z (PI) signals	N° m/z (PI) signals in the model	N° m/z (NI) signals	N° m/z (NI) signals in the model
BB vs CG	0.9473	0.8632	2	2211	10	465	4
BB vs DG	0.8756	0.8094	1	2334	35	492	4
TN vs CG	0.9839	0.8705	2	2207	56	470	14
TN vs DG	0.9814	0.8107	2	2315	143	496	40
PL vs CG	0.9904	0.8639	2	2315	51	495	15
PL vs DG	0.9905	0.9611	3	2407	9	517	1
CG vs DG	0.9652	0.7358	2	2347	150	487	35

Table 4.8: PLS-DA classification models obtained comparing the EVOO samples stored in clear glass (CG) and dark glass (DG) with those stored in bag-in-box (BB), tin (LA) and plastic (PL).

The classification models offer a complete picture of the information present in the EVOO samples, as demonstrated by the fitting correlation coefficients (R^2), all greater than 0.87; furthermore, all the models are characterized by a high predictive ability, since all the cross validated correlation coefficients (R^2_{CV}) are greater than 0.73. A great reduction of the number of significant variables was obtained by using PLS-DA, making the data easier to interpret.

As an example, hereafter the comparison between clear glass (CG) and plastic (PL) will be discussed. The plot reported in Figure 4.10 shows the predicted (in red) and fitting (in blue) values of the class score, showing how the samples stored in CG and PL are all correctly grouped into two classes, as described by the R^2 and R^2_{CV} values of 0.9904 and 0.8639, respectively.
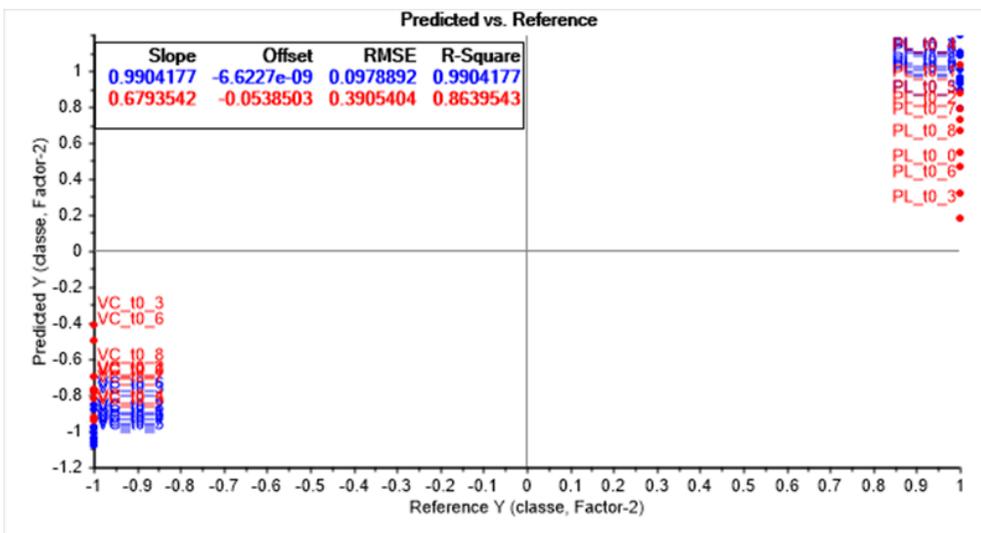
Figure 4.10: predicted (in red) and observed (in blue) objects for the samples stored in CG and PL.

The score plot (Fig. 4.11) represents the objects in the area described by the first two latent variables that together represent 99% of the information related to the type of packaging material (Y variable) and 28% of that related to the acquired signals (X matrix). The objects are clearly separated in the two classes, as demonstrated by the goodness of the model (Tab. 4.8).
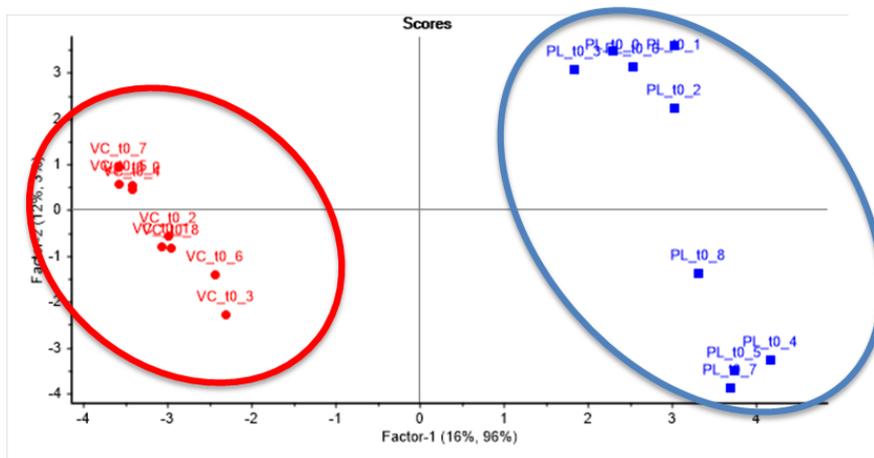

Figure 4.11: score plot of the two first latent variables that allows the correct classification of EVOO samples stored in CG (*red ellipse*) and in PL (*blue ellipse*).

The loading plot (Fig. 4.12) allows the identification of the variables that mostly contribute to the differentiation of the EVOO samples stored in CG and PL. The variables circled in blue show higher values for EVOO samples stored in PL and lower values for those stored in CG; the variables circled in red show the opposite trend.
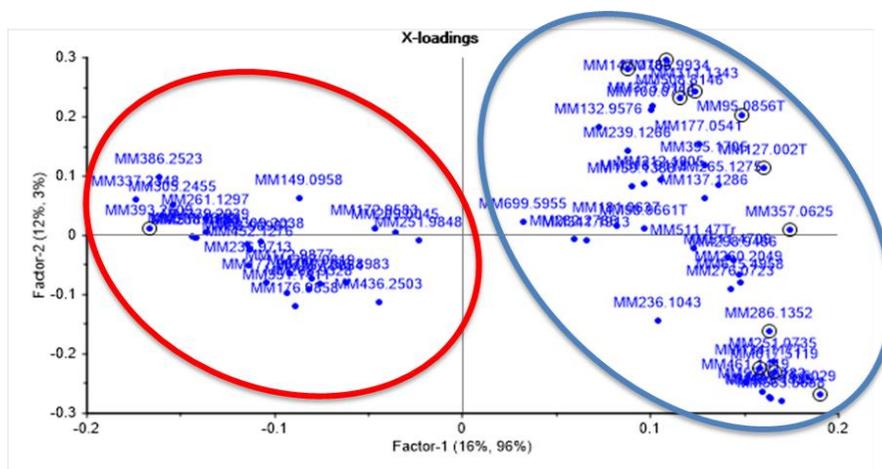

Figure 4.12: loading plot of the first two latent variables that allows the correct separation of the samples stored in CG and PL.

Through the evaluation of the relationships between the intensities of the signals considered as significant by the chemometric analysis, it was possible to identify a set of *m/z* characteristic of each of the packaging material considered.

4.4 Conclusions

A new UHPLC-MS/MS method for the identification of eight polyphenols and pantothenic acid was developed and validated by evaluating its linearity range, correlation coefficients, LODs, LOQs, inter- and intra- day precision. No matrix effect was evidenced for any of the investigated analytes. The right combination of stationary phase and mobile phase provided an excellent sensitivity. Furthermore, the developed methodology is easy to reproduce and shows high recovery values for all the target compounds. The method was applied to the 115 investigated samples in order to assess which packaging material is the most suitable for EVOO

beneficial properties preservation, evaluating the PPs concentration trend in all the containers. Dark glass demonstrated to be the best packaging material among the considered ones: it better protects EVOO from external agents as indicated by the slow decrease of quercetin concentration (trend) both in containers monthly opened and immediately analyzed and in those opened and monthly analyzed (conservation and ageing studies).

Each sample was further investigated by the non-target approach: the unselective conditions adopted for 5600^+ TripleTOF^{TM} analyses allowed the acquisition of the complete fingerprint of each sample, as the instrument was working in Data Independent Scan Mode. Direct comparison between the EVOO samples and the control allowed the identification of oleocanthal in all the investigated samples. Such compound, in fact, is the responsible for the particular burning sensation occurring in throat while consuming EVOO. Given the enormous quantity of data available, chemometric techniques were adopted in order to reduce the redundant information, and to highlight the information only ascribable to the different material adopted for storing the EVOO. Principal Component Analysis did not allow the identification of compounds able to characterize the different packaging materials. Conversely, performing Partial Least Square-Discriminant Analysis led to the identification of about 1700 *m/z* signals (considering both positive and negative signals and all the comparisons reported in tables 4.7 and 4.8 together) that can differentiate the EVOO samples stored in the different packaging materials.

References

1. Mazzotti F, Benabdelkamel, Di Donna L, Maiuolo L, Napoli A, Sindona G (2012) Food Chem 135: 1006-1010.

2. Garcìa-Villalba R, Carrasco-Pancorbo A, Oliveras-Ferraros C, Vàzquez-Martìn A, Menéndez J A, Segura-Carretero A, Fernàndez-Gutiérrez A (2010) J Pharm Biomed Anal 51: 416-429.

3. Trìpoli E, Giammanco M, Tabacchi G, Di Majo D, Giammanco S, La Guardia M (2005) Nutr Res Rev 18: 98-112.

4. Owen R R, Giacosa A, Hull W E, Haubner R, Spiegelhalder B, Bartsch H (2000) Lancet Oncol 36: 1235-1247.

5. Menéndez J A, Vàzquez-Martin A, Olivares-Ferraros C, Garcìa-Villalba R, Carrasco-Pancorbo A, Fernàndez-Gutierréz A, Segura-Carretero A (2009) Int J Oncol 34: 43-51.

6. Menéndez J A, Vàzquez-Martin A, Olivares-Ferraros C, Garcìa-Villalba R, Carrasco-Pancorbo A, Fernàndez-Gutierréz A, Segura-Carretero A (2008) Int J Mol Med 22: 433-439.

7. Cicerale S, Conlan X A, Sinclair A J, Keast R S J (2009) Crit Rev Food Sci Nutr 49: 218-236.

8. Menéndez J A, Vàzquez-Martin A, Garcìa-Villalba R, Carrasco-Pancorbo A, Oliveras-Ferraros C, Fernàndez-Gutierréz A, Segura-Carretero A (2008) BMC Cancer 18: 377-400.

9. Amati L, Campanella L, Dragone R (2008) J Agric Food Chem 56: 8287-8295.

10. Rizzo V, Torri L, Liciardello F, Piergiovanni L, Muratore G (2014) Packag Technol Sci 27: 437-448.

11. Fabiani R, Rosignoli P, Bartolomeo A, Fucelli R (2008) J Nutr 138: 1411-1416.

12. Servili M, Montedoro G (2002) Eur J Lip Sci Technol 104: 602-613.

13. Asparicio R, Roda L, Albi M A, Gutierrez F (1999) J Agric Food Chem 47: 4150-4155.

14. Lavelli V, Bondesan L (2005) J Agric Food Chem 53: 1102-1107.

15. Baccouri O, Cerretani L, Bendini A, Lercker G, Zarrouk M, Ben Miled D D (2008) Riv Ital Sostanze Gr 85: 189-195.

16. Cerretani L, Salvador M D, Bendini A, Fregapane G (2008) Chemosens Percept 1: 258-267.

17. Gutiérrez-Rosales F, Rios J J, Gómez-Rey M L (2003) J Agric Food Chem 51: 6021-6025.

18. Carrasco-Pancorbo A, Cerretani L, Bendini A, Segura-Carretero A, Del Carlo M, Gallina-Toschi T, Lercker G, Compagnone D, Fernández-Gutiérrez A (2005) J Agric Food Chem 53: 8918-8925.

19. Commission Regulation (EC) No 432/2012 of 16 May 2012, Off J Eur Commun (2012) 136: 1-40.

20. Berzas Nevado J J, Castañeda Peñalvo G, Rodríguez Robledo V, Vargas Martinez G (2009) Talanta 79: 1238-1246.

21. Armaforte E, Mancebo-Campos V, Bendini A, Desamparados Salvador M, Fregapane G, Cerretani L (2007) J Sep Sci 30: 2401-2406.

22. Grumetto L, Montesano D, Seccia S, Albrizio S, Barbato F (2008) J Agric Food Chem 56: 10633-10637.

23. Errico S, Bianco M, Mita L, Migliaccio M, Rossi S, Nicolucci C, Menale C, Portaccio M, Gallo P, Mita D G, Diano N (2014) Food Chem 160: 157-164.

24. Sungur S, Köroğlu M, Özkan A (2014) Food Chem 142: 87-91.

25. Perring L, Basic-Dvorzak M (2002) Anal Bioanal Chem 374: 235-243.

26. Pasias I N, Papageorgiou V, Thomaidis N S, Proestos C (2012) Food Anal Methods 5: 835-840.

27. Lavelli V, Fregapane G, Desamparados S M (2006) J Agric Food Chem 54: 3002-3007.

28. Gómez-Alonso S, Mancebo-Campos V, Desamparados Salvador M, Fregapane G (2007) Food Chem 100: 36-42.

29. Riahi J, Marzouk B (2000) Riv Ital Sostanze Gr 77: 25-30.

30. Mendez A I, Falqué E (2007) Food Control 18: 521-529.

31. Okogeri O, Tasioula-Margari M (2002) J Agric Food Chem 50: 1077-1080.

32. Bonoli M, Bendini A, Cerretani L, Lercker G, Toschi T G (2004) J Agric Food Chem 52: 7026-7032.

33. DeLa Torre-Carbot K, Jauregui O, Gimeno E, Castellote A I, Lamuela-Raventos R M, Lopez-Sabater M C (2005) J Agric Food Chem 53: 4331-4340.

34. Pizarro M L, Becerra M, Sayago A, Beltrán M, Beltrán R (2013) Food Anal Meth 6: 123-132.

35. Mateos R, Espartero J L, Trujillo M, Rios J J, Leon-Camacho M, Alcudia F, Cert A (2001) J Agric Food Chem 49: 2185-2192.

36. Selvaggini R, Servill M, Urbani S, Esposto S, Taticchi A, Montedoro G (2006) J Agric Food Chem 54: 2832-2838.

37. Rios J J, Gil M J, Gutierrez-Rosales F (2005) J Chromatogr A 1093: 167-176.

38. Garcia-Villalba R, Pacchiarotta T, Carrasco-Pancorbo A, Segura-Carretero A, Fernández-Gutiérrez A, Deelder A M, Mayboroda O A (2011) J Chromatogr A 1218: 959-971.

39. Suàrez M, Macià A, Romero M-P, Motilva M J (2008) J Chromatogr A 1014: 90-99.

40. Fu S, Segura-Carretero A, Arráez-Román D, Menendez J A, DeLa Torre A, Fernández-Gutiérrez A (2009) J Agric Food Chem 57: 11140-11147.

41. Gómez-Caravaca A M, Carrasco-Pancorbo A, Cañabate-Díaz B, Segura-Carretero A, Fernández-Gutiérrez A (2005) Electrophoresis 26: 3538-3551.

42. Carrasco-Pancorbo A, Arráez-Román D, Segura-Carretero A, Fernández-Gutiérrez A (2006) Electrophoresis 27: 2182-2196.

43. Carrasco-Pancorbo A, Gómez-Caravaca A M, Cerretani L, Bendini A, Segura-Carretero A, Fernández-Gutiérrez A (2006) J Agric Food Chem 54: 7984-7991.

44. Garcia-Villalba R, Carrasco-Pancorbo A, Vázquez-Martin A, Olivers-Ferraros C, Menéndez J A, Segura-Carretero A, Fernández-Gutiérrez A (2009) Electrophoresis 30: 2688-2701.

45. Gómez-Caravaca A M, Carrasco-Pancorbo A, Segura-Carretero A, Fernández-Gutiérrez A (2009) Electrophoresis 30: 3099-3109.

46. Aturky Z, Fanali S, D'Orazio G, Rocco A, Rosati C (2008) Electrophoresis 29: 1643-1650.

47. Becerra-Herrera M, Sanchez-Astudillo M, Beltran R, Sayago A (2014) Food Sci Technol 57: 49-57.

48. Gilbert-López B, Valencia-Reyes Z L, Yufra-Picardo V M, García-Reyes J F, Ramos-Martos N, Molina-Díaz A (2014) Food Anal Method 7: 1824-1833.

49. Hrncirik K, Fritsche S (2004) Eur J Lipid Sci Technol 106: 540-549.

50. Carrasco-Pancorbo A, Neusüβ C, Pelzing M, Segura-Carretero A, Fernández-Gutiérrez A (2007) Electrophoresis 28: 806-821.

51. Carrasco-Pancorbo A, Cerretani L, Bendini A, Segura-Carretero A, Fernández-Gutiérrez A (2005) J Sep Sci 28: 837-851.

52. Alarcón Flores M I, Romero-Gonzalez R, Garrido Frenich A, Martinez Vidal J L (2012) Food Chem 134: 2465-2472.

53. Saitta M, Salvo F, Di Bella G, Dugo G, La Torre L G (2009) Food Chem 112: 525-532.

54. García-Villalba R, Carrasco-Pancorbo A, Zurek G, Behrens M, Bäβmann C, Segura-Carretero A, Fernández-Gutiérrez A (2010) J Sep Sci 33: 2069-2078.

55. Perri E, Raffelli A, Sindona G (1999) J Agric Food Chem 47: 4156-4160.

56. Godoy-Caballero M, Galeoano-Diaz T, Acedo-Valenzuela M I (2012) J Sep Sci 35: 3529-3539.

57. delos Angeles De Fernandez M, Soto Vargas V C, Silva M F (2014) J Am Oil Chem Soc 91: 2021-2033.

58. Bianco A, Buiarelli F, Cartoni G, Coccioli F, Jasionowska R, Margherita P (2003) J Sep Sci 26: 417-424.

59. Armaforte E, Mancebo-Campos V, Bendini A, Salvador M D, Fregapane G, Cerretano L (2007) J Sep Sci 30: 2402-2406.

60. Carrasco-Pancorbo A, Gómez-Caravaca A M, Cerretani L, Bendini A, Segura-Carretero A, Fernández-Gutiérrez A (2006) J Sep Sci 29: 2221-2233.

61. Gallina-Toschi T, Cerretani L, Bendini A, Bonoli-Carbognin M, Lercker G (2005) J Sep Sci 28: 859-870.

62. Carrasco-Pancorbo A, Segura-Carretero A, Fernández-Gutiérrez A (2005) J Sep Sci 28: 925-934.

63. Capriotti A L, Cavaliere C, Crescenzi C, Foglia P, Nescatelli R, Samperi R, Laganà A (2014) Food Chem 158: 392-400.

64. Gosetti F, Bolfi B, Manfredi M, Calabrese G, Marengo E (2015) J Sep Sci 38: 3130-3136.

65. Eurachem Guide: The Fitness for Purpose of Analytical Methods-A Laboratory Guide to Method Validation and Related Topics, 2nd edition, Magnusson B, Örnemark U (Eds), 2014.

66. Ryan D, Robards K (1998) Analyst 123: 31-44.

Chapter 5

Use of Data Independent Acquisition mode for identification and determination of target and unknown pollutants species in paddy water samples

5.1 Introduction

Weeds proliferation in paddy fields extremely affects rice production. To protect their crops, farmers widely employ combination of herbicides, since they provide a pre- and post-emergence activity due to the inhibition of plant enzymes production. Sopeña et *al*. [1] reported that worldwide consumption of herbicides represents 47.5% of the 2 million tons of pesticides consumed each year. Industries efforts are addressed to the design of new formulations characterized by high efficiency at low doses and low persistence in fields in favor of the minimum environmental impact possible. However, their intrinsic features (i.e. molecular weight, polarity, solubility and partition coefficient octanol/water, K_{ow}) make pesticides able to spread in water, to leach into the soil or to evaporate, making their environmental impact evaluation of utmost importance. Furthermore, since rice cultivation is characterized by flooded shallow environments, herbicides come into contact with paddy scenery through the draining of water into channels and, thanks to their mobility, they can contaminate areas not directly subjected to phytosanitary treatments. Water resources contamination, in fact, is one of the major problems to deal with in order to ensure conservation and sustainability of rice environment. Pesticides represent 20 out of 45 hazardous substances identified in water by the European Water Framework Directive [2]. There are other substances able to alter

the environmental ecosystems: they represent the class of "*emerging contaminants*". These are structurally different one from each other, consisting of a heterogeneous group of molecules: hormones, pharmaceuticals, endocrine disrupting compounds, per-fluorinated compounds, flame retardants, personal care products, and transformation products of the above listed substances. In particular, there is a growing interest in contaminants transformation products, since they could be more toxic and present at higher concentrations than their precursor [3-5]. It is easy to understand that the problems related to the use of herbicides are different: damages to non-target flora, including the rice one (possible residues in fruits), presence of residues and degradation products in different environmental compartments, leading to possible toxicity not only for aquatic organisms, but also for humans; furthermore an intensive use of a specific chemical could induce weed resistance.

Objects of this study are three herbicides widely used in rice crops: Imazamox (IMA, commercialized as Beyond^{TM}, containing 3.7% w/w of the active principle IMA; 0.04 Kg of Beyond^{TM} for 200-300 L/ha), Oxadiazon (OXA, commercialized as Ronstar^{TM}, containing 34.86% w/w of the active principle OXA; 0.38 Kg of Ronstar^{TM} for 500 L/ha) and Profoxydim (PRO, commercialized as Aura^{TM}, containing 20.6% w/w of the active principle PRO; 0.20 Kg of Aura^{TM} for 200-300 L/ha), that act on acetolactate synthase, protoporfhyrinogen oxidase and acetyl CoA carboxylase of weed plants, respectively. Figure 5.1 shows the three investigated herbicides.

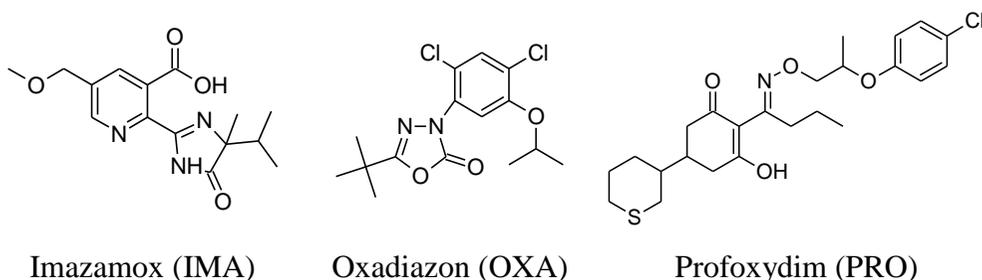Imazamox (IMA)   Oxadiazon (OXA)   Profoxydim (PRO)

Figure 5.1: chemical structures of Imazamox (IMA), Oxadiazon (OXA) and Profoxydim (PRO).

As previously mentioned, the active principles can be transformed into new species (because of the degradation reactions), since they are susceptible to natural modifications ascribable to the effect of sunlight, variations of pH and temperature or interactions with other species naturally present in water. The impact of the above mentioned causes is strictly related to the structure of the compounds. Carboxylic acids, alcohols and dealkylated derivatives are the main OXA degradation products in rice [6-10]. PRO, on the contrary, leads to oxazoles and isoxazoles that are compounds more soluble in water and show less soil sorption attitude compared to their precursor. Previous studies demonstrated that IMA can undergo to photodegradation processes when dissolved in ultrapure water, generating degradation products due to the replacement of the acidic and methoxy groups by a hydroxyl one in the pyridine moiety, and the breaking of the C-N and C-O bonds in the imidazole moiety [11].

This study, on the contrary, is focused on the natural photodegradation of the three herbicides, that can take place when they are dispersed into the water that floods the paddy fields. For this aim, the herbicides were dissolved into paddy water, in order to reproduce all the possible interactions that can take place between the compounds naturally present in water. In fact, it was demonstrated that the medium in which the contaminants are dissolved can influence their degradation processes, leading to different degradation products [3-5, 12]. Given the high complexity of the matrix, it is necessary to develop analytical methods able to identify all the

different contaminants. In literature, methods are developed to meet the requirements of high sensitivity necessary to quantify pesticides, whose allowed concentrations are extremely low, in the order of ng L^{-1} [13, 14]. Gas and liquid chromatography techniques are the most suitable separative ones. Even if GC is widely employed [7, 15-18], LC fits best the low volatility, thermolability and polar features of the herbicides [11].

The aim of this work is the evaluation of water quality of the considered paddy fields through the identification not only of the three selected pesticides, but also of other emerging contaminants, such as their photodegradation products and other non-target compounds. For this purpose, a new sensitive micro liquid chromatography method coupled with tandem high-resolution mass spectrometry was developed and validated.

A xenon lamp was employed to degrade IMA, OXA and PRO, simulating the incidence of solar radiation during submersion of the paddy fields. The three herbicides were determined through the target approach, while their degradation products and other emerging contaminants were identified thanks to the non-target one, exploiting the Data Independent Acquisition mode provided by the employed high-resolution mass spectrometer. The method was applied to groundwater and surface water samples collected from paddy fields in the Padana plain in two agricultural seasons. The obtained data were used to evaluate the behavior of pesticides and to detect possible degradation products that might persist in water and diffuse in the environment, trying to offer a comprehensive study of paddy field water ecosystem.

5.2 Experimental

5.2.1 Reagents

Reagent	Purity	Vendor
Methanol	LC-MS Ultra CHROMASOLV, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Ammonium acetate	> 99.995 %	Sigma–Aldrich (Milwaukee, WI, USA)
Formic acid	Eluent additive, purity > 99.9 %	Sigma–Aldrich (Milwaukee, WI, USA)
Oxadiazon	Analytical standard grade	Sigma–Aldrich (Milwaukee, WI, USA)
Imazamox	Analytical standard grade	Sigma–Aldrich (Milwaukee, WI, USA)
Profoxydim lithium salt	Analytical standard grade	Sigma–Aldrich (Milwaukee, WI, USA)
Water	For mass spectrometry	Fluka (Buchs, Switzerland)

Table 5.1: list of the used reagents.

5.2.2 Apparatus

In order to simulate sunlight irradiation, a CoFoMeGra solar box 3000e (Milan, Italy) equipped with a xenon lamp was adopted. The analyses were performed using a micro LC system (Eksigent Technologies, Dublin, USA) that included a micro LC200 pump with a flow module of 5-50 μL and a CTC programmable array logic autosampler with a Peltier unit (1-45 °C). The system was coupled with a TripleTOF^{TM} 5600$^+$ system (Sciex, Concord, Canada) equipped with a DuoSpray^{TM} ion source and calibrant delivery system (CDS). PeakView^{TM} 2.1 (Sciex, Concord, Canada) was adopted to investigate accurate mass spectral data, while MasterView^{TM} 1.1 (Sciex, Concord, Canada) was used to compare the intensities of mass spectral data among different samples. MultiQuant^{TM} 3.0 (Sciex, Concord, Canada) was used to process and to quantify the herbicides in water samples.

5.2.3 Stock solutions and sample preparation

The three herbicides stock solutions were prepared in 0.2 μm PTFE filtered (VWR International, Darmstadt, Germany) paddy water (pH=8.7) and preserved at 4 °C in dark glass vials: in particular, PRO solution was prepared at its maximum solubility in water at 26.5 mg L^{-1} as reported by the EPA agency [19], the IMA solution was prepared at 30 mg L^{-1}, in order to be comparable with that of PRO. OXA maximum water solubility was evaluated in our laboratory since literature data were contrasting, reporting different values [21, 22]. A calibration plot was built with standard solutions of OXA dissolved in methanol, considering the concentration range between 0.025 mg L^{-1} and 5.0 mg L^{-1}; the obtained experimental water solubility value was 0.1 mg L^{-1}.

Water samples were collected from an irrigation district representative of the Ticino-Sesia paddy farmland in Northern Italy (Fig. 5.2). Four different fields were considered: 1 and 2 were cultivated with dry-seeding and delayed flooding technique, while 3 and 4 with the wet seeding and continuous flooding one. In order to investigate the contaminants present in the water of the selected area, 18 sampling points were considered: 8 points for the surface water (labeled as "S" from "S1" to "S8"), 2 for underground water samples collected through piezometers (P1 out of the farm and P2 inside the farm) and 8 from the porous cups (the 4 cups labeled as "Ca" refer to 25 cm soil depth and the 4 cups labeled as "Cb" refer to the 60 cm ones). All the water samples were collected every two weeks during the two agricultural seasons (first agricultural season: from April to September 2015; second agricultural season: from April to October 2016); sampling frequency became daily (once a day for a week after every fields treatments with any herbicide) during the period in which fields were weeded.

The samples were collected, filtrated on a 0.2 μm PTFE filter (VWR International, Darmstadt, Germany) and immediately subjected to LC-MS/MS analysis.
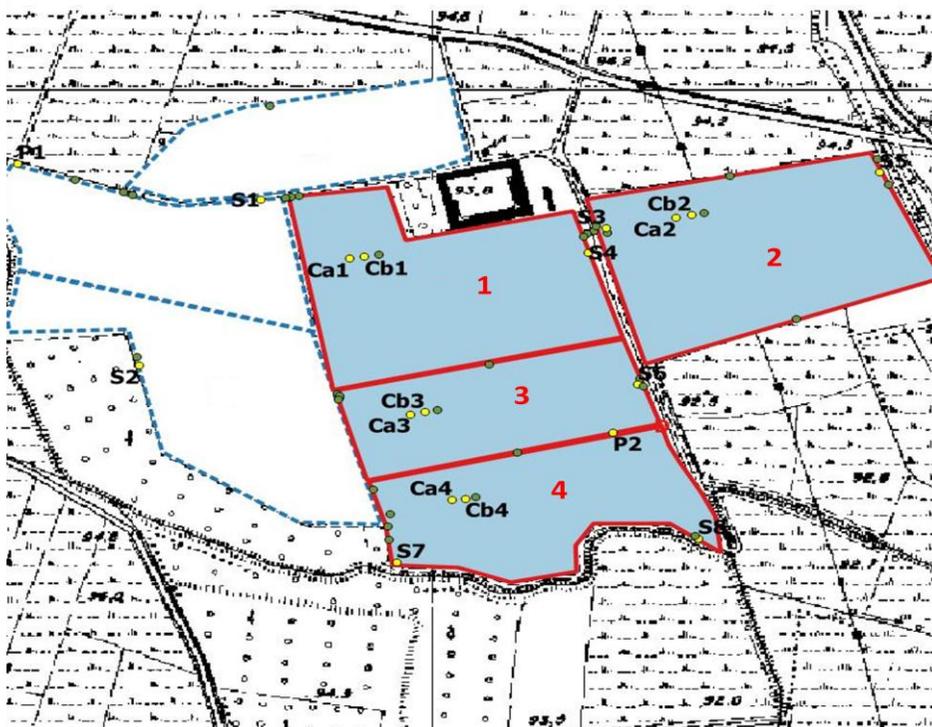
Figure 5.2: Paddy fields and sampling points object of the present study.

5.2.4 Chromatographic and mass spectrometric conditions

The selected stationary phase was a HALO C18 column (0.5 mm x 100 mm, 2.7 μm; Eksigent Technologies, Dublin, USA), while the mobile phase was constituted by a mixture of ammonium acetate 1.0 mM and 0.1 % (v/v) formic acid in water (A) and in methanol (B), eluting at a flow-rate of 25 μL min^{-1}, according to the gradient reported in Table 5.2.

166

Time (min)	% A	% B
0.0	90.0	10.0
0.5	90.0	10.0
5.0	5.0	95.5
7.0	5.0	95.5
7.1	90.0	10.0
10.0	90.0	10.0

Table 5.2: elution gradient of ammonium acetate 1.0 mM and 0.1% (v/v) formic acid in water (A) and in methanol (B).

The injection volume was set at 150 nL, while the column oven temperature was 45 °C. The DuoSpray[TM] ion source worked in positive ion mode (PI). The instrumental conditions were set as follows: curtain gas (N$_2$) at 25 psig, nebulizer gas GS1 and GS2 at 25 and 35 psig respectively, desolvation temperature at 400 °C and ion spray floating voltage at +5500 V. During time-of-flight (TOF) MS experiments, the declustering potential (DP) and the collision energy (CE) were respectively set at +100 and +10 V, while during the TOF-MS/MS experiments CE was set at +40 V. SWATH[TM] (Data Independent Acquisition) analysis was exploited to perform non-target analyses, providing the complete fingerprint of each analyzed sample: such acquisition, in fact, generates an exhaustive ion map of the fragments from all the detectable precursor masses (Paragraph 1.4.5). SWATH[TM] distributes all the ions in subsequent windows: all the precursor ions are transferred to the collision cell where they are fragmented to generate the corresponding product ions of the MS/MS spectrum, and then traced back to the precursor ion through a deconvolution software. To this aim, the TOF scan in a mass range of 100-900 Da (accumulation time 100 ms) is followed by MS/MS scans of the precursor ions in the range of 100-900 Da; this range was divided into 32 windows each 25 Da wide, with an accumulation time of 20 ms for each window. SWATH[TM] approach ensures that all the possible information related to

any of the precursor ion is acquired, allowing a complete and accurate screening of compounds present at low concentrations.

5.2.5 Solar box irradiation conditions and hydrolysis studies

The three herbicide solutions (20.0 mg L^{-1} for IMA and PRO, while 0.1 mg L^{-1} for OXA) were introduced into quartz cells (volume of 14 mL) and subjected to solar box irradiation for 24 hours. In order to mimic the environmental conditions in which the paddy fields are submerged into water in the flooding season, the solar box xenon lamp irradiation intensity was set at 800 W m^{-2} and temperature at 35 °C; a UV glass filter was used to simulate the outdoor conditions. Such parameters were set in agreement with the information provided by the meteorological station installed in field 3.

Other three solutions of the considered herbicides were prepared at the same concentration and preserved in the dark for 30 days, in order to evaluate possible hydrolysis reactions (without any contribution from the sunlight irradiation).

All the experiments were replicated three times.

5.3 Results and discussion

5.3.1 Development of UHPLC-MS/MS method

A preliminary MS/MS characterization study of the three herbicides in positive ion mode (PI) was performed in order to collect information useful for the identification of the degradation products, that could be formed during the irradiation processes. Each herbicide standard solution (0.100 mg mL^{-1} in methanol) was infused through the syringe pump at 10.00 μL min^{-1} into the mass spectrometer. The most intense signals corresponding to the [M+H]$^{+}$ species observed for the herbicides, their two most intense product ions, that were used for their quantification and their corresponding structures, are reported in Table 5.3.

Analyte	[M+H]$^+$	Product ions (*m/z*)	t$_R$ (min)	Proposed product ion structure
IMA	306.1	193.0612	3.2	
		163.0497		
OXA	345.1	176.9504	6.1	
		219.9559		
PRO	466.2	180.1025	5.9	
		280.1381		

Table 5.3: [M+H]$^+$ species of the three pesticides, their most intense product ions, their structures and retention times (t$_R$).
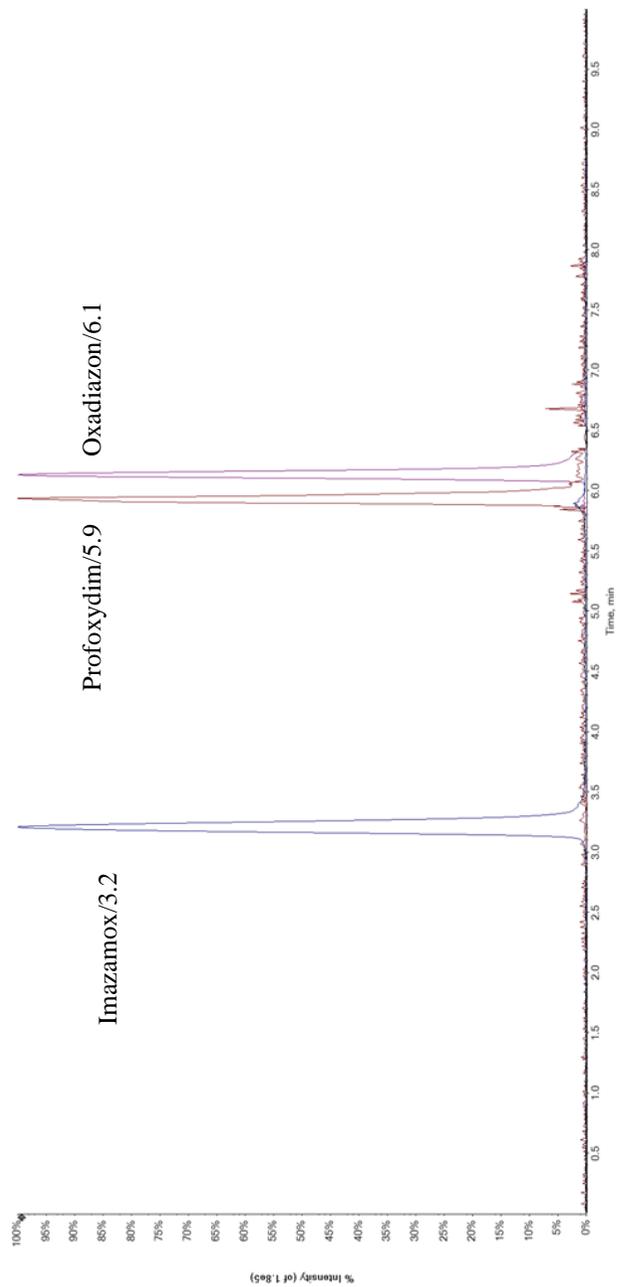
Figure 5.3: chromatographic separation of the mixture of analytes (250.0 µg L^{-1} each) in the best experimental conditions.

Figure 5.3 shows the chromatographic separation of the three herbicides at 250.0 µg L^{-1} each, prepared in paddy water.

5.3.2 Photodegradation and hydrolysis studies

OXA, IMA and PRO solutions were irradiated for 24 hours since they are supposed to act immediately after the application and they are not expected to persist for a long time in the flooded paddy environment. At prefixed times, aliquots of the solutions were collected and subjected to LC-MS/MS analyses. With the proceeding of the degradation process, the chromatographic peak intensities of the signals related to the three active principles decreased, while new peaks ascribable to new formed species arose.

The degradation kinetics were evaluated plotting *lnC* (being *C* the concentration of the analyte) as a function of the irradiation exposure: the obtained linear fit indicated a first order kinetic degradation process for all the analytes (Fig. 5.4). The calculated half-lives (*t$_{1/2}$*) for IMA, OXA and PRO were respectively 63 min(h=1.05), 193 min (h=3.22) and 322 min (h=5.37).

t$_{1/2}$= 63min (h = 1.05)   t$_{1/2}$ = 193min (h = 3.22)   t$_{1/2}$ = 322min (h = 5.37)
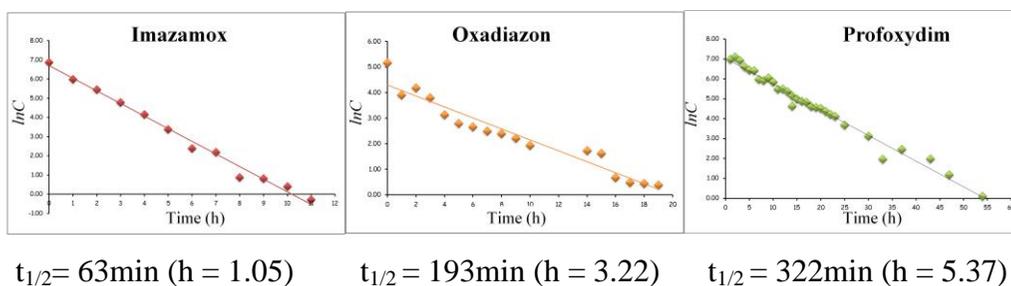
Figure 5.4: plot of *lnC* as a function of irradiation exposure time.

In order to evaluate the possible reactions due to hydrolysis, the solutions of IMA, OXA and PRO, were also preserved in the dark for 30 days. At prefixed times, aliquots of each solution were collected and subjected to LC-MS/MS analysis. The

results demonstrated that no hydrolysis reactions were taking place, as the initial concentrations of the three herbicides were not statistically different from the final ones and did not change significantly along the time.

5.3.3 Method validation

For each herbicide a calibration plot reporting the peak area of the most intense transition (*y*) versus the standard concentration (*x*) was built, considering 5 concentration levels, ranging between the LOQ value and 250.0 µg L^{-1} (each concentration level was replicated three times). The standard solutions were injected in a randomized order to avoid possible memory effects. A 1/x weighted linear regression fit was adopted to interpolate the analytes concentrations. The three calibration plots were characterized by a good linearity and R^2 correlation coefficients values, all greater than 0.9989 (Tab. 5.4).

The Limit of Detection (LOD) was calculated as the lowest detectable concentration of the analyte that can be distinguished from the background, calculated as:

$$LOD = S_{blank} + 3s_{blank}$$

being S_{blank} the signal (peak area) equal to the average background and s_{blank} its standard deviation. The Limit of Quantification (LOQ) represents the lowest concentration quantifiable in the linearity range, calculated as:

$$LOQ = S_{blank} + 10s_{blank}$$

Intra- and inter-day precision of the retention times and concentrations were evaluated through the analysis of a standard mixture of the three herbicides (250.0 µg L^{-1} each), five replicates per day for a week. The obtained results demonstrated

that the intra-day relative standard deviation (RSD%) of the concentration was lower than 4.7% and the inter-day RSD% was lower than 6.1%. Intra- and inter-day RSD% calculated on the retention time were lower than 0.57% and 0.72%, respectively (Tab. 5.4). the matrix effect (*ME*) was evaluated performing a *t*-test at 95% confidence level on the ratio between the slope of the external calibration plot (slope$_{ext}$) and that of the standard addition plot (slope$_{add}$) obtained by spiking the analyte mixture (10.0, 100.0 and 250.0 μg L^{-1}) in a blank paddy water solution. *ME* was calculated as percentage [5]:

$$ME\% = \frac{slope_{add}}{slope_{ext}} \cdot 100 - 100$$

ME results equal to 0 for all the considered analytes, so no relevant *ME* was found.

Analyte	Linearity range (pg mL^{-1})	R^2	LOD (pg mL^{-1})	LOQ (pg mL^{-1})	RSD % intra day (n=5) t$_R$	RSD % inter day (n=35) t$_R$	RSD % intra day (n=5) (conc.)	RSD % inter day (n=35) (conc.)
IMA	13-250000.0	0.9998	4	13	0.57	0.72	4.7	6.1
OXA	17-250000.0	0.9995	5	17	0.16	0.25	2.4	2.7
PRO	13-250000.0	0.9989	4	13	0.25	0.46	2.0	2.3

Table 5.4: linearity range, regression coefficient (R^2), LOD, LOQ, intra- and inter-day RSD (%) of retention times and of concentrations.

5.3.4 Identification of photodegradation products

In order to identify the degradation products formed during the photoirradiation of the three herbicides, a "direct comparison" between the irradiated samples (degraded) and the non-irradiated (control) ones was first performed by the software in a semi-automated mode. Before performing the comparison, the following parameters were set:

- Do not calculate details for XIC intensity < 200 counts;
- The signal to noise ratio < 10
- The XIC width (Da) = 0.02
- Retention time width (min) = 0.5
- Threshold (cps) = 1000
- Threshold (ratio of control) = 5
- Minimum retention time = 0.3
- Maximum retention time = 9

The threshold value set at 1000 counts ensures that the attention is focused on the more abundant photoproducts that can be generated because of the degradation of the three herbicides in paddy water due to solar irradiation. A further constraint was defined by the intensity ratio between the irradiated and non-irradiated samples: in this study, this ratio was set at 5. Only the peaks that met all these requirements were automatically included in a list by the software MasterView[TM] and then extracted from the total ion current (TOF scan).

The irradiation of the paddy water solutions containing the three herbicides, generated 2 different photoproducts for IMA (P1_IMA and P2_IMA), 3 photoproducts for PRO (P1_PRO, P2_PRO and P3_PRO), while no photodegradation product was identified in OXA solutions.

Previous OXA degradation products investigations were performed using GC-MS or electroanalytical methods [6-10, 16-18, 22-26]. These approaches led to the identification of degradation products, whose presence was not observed in our LC-MS data.

The chemical structures elucidation of IMA and PRO degradation products was based on the elemental composition within a mass tolerance of 1.1 ppm, the isotopic cluster abundance (tolerance of 10%), the number of rings and double bonds (RDBs), the use of the high-accurate and high-resolution MS/MS acquired spectra, from which it was possible to obtain the elemental composition of the product ions (considering an average mass tolerance lower than 7.0 ppm), their retention time plausibility in relation with their calculated Log*P*. In particular, the information obtained through the MS/MS fragmentation are fundamental for the interpretation of the precursor molecular structures. Once the chemical structures of the degradation products were proposed, they were drawn in PeakView[TM] where their experimental MS/MS spectra can be compared with that simulated *in silico* for the same structures, offering a matching percentage as a result. All the proposed structures gave positive matching, all greater than 92%.

Table 5.5 reports the elemental composition, the accurate *m/z*, the number of rings and double bonds (RDB), the mass error (ppm) and the average MS/MS error (ppm) calculated through the number of MS/MS product ions for each of the photodegradation products reported in Figure 5.5. Figure 5.5 reports the proposed structures of the degradation products of IMA and PRO.

[M+H]$^{+}$ species	Precursor herbicide	Molecular formula	Accurate m/z	t_R (min)	RDB	MS error (ppm)	MS/MS error (ppm)
P1_IMA	IMA	$C_9H_9NO_4$	196.0604	4.5	6	-0.2	15.1 (7)
P2_IMA	IMA	$C_{14}H_{18}N_2O_3$	263.1390	3.4	7	1.1	2.1 (12)
P1_PRO	PRO	$C_{15}H_{23}NO_2S$	282.1522	5.7	5	0.6	5.7 (7)
P2_PRO	PRO	$C_{24}H_{32}ClNO_5S$	482.1762	6.0	9	1.1	2.0 (6)
P3_PRO	PRO	$C_{15}H_{21}NO_2S$	280.1366	5.8	6	0.4	3.7 (16)

Table 5.5: origin of the degradation product, elemental composition, accurate m/z, retention time, number of rings and double bonds (RDBs), mass error, and intensity weighted average error for the tandem mass spectrometry (MS/MS) data calculated through the number of the product ions reported in parentheses.
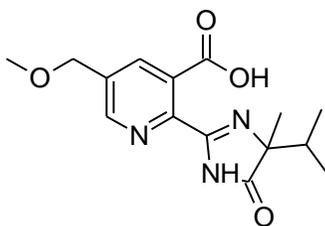
IMA
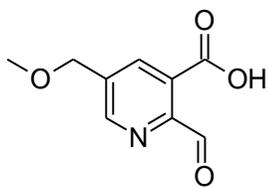
P1_IMA
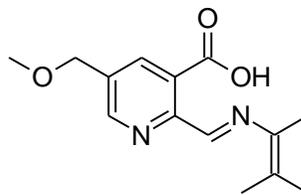
P2_IMA
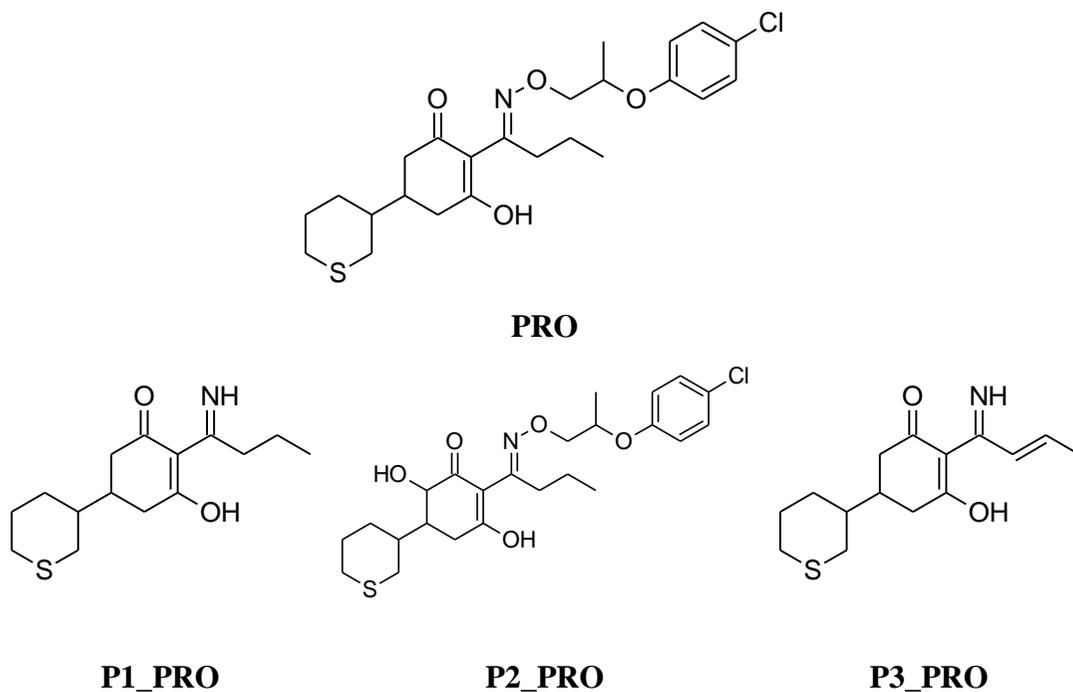
Figure 5.5: proposed chemicals structures of the IMA and PRO photodegradation products.

Both the IMA degradation products maintain the portion of the molecule characterized by the substituted pyridine ring; in particular, P1_IMA is formed by the opening and subsequent loss of the imidazole ring, while P2_IMA by the opening of the same ring and loss of the amidic group and consequent rearranging of the alkyl chain.

PRO degradation products are associated with the preservation of the thiane ring, but differ for the loss of the O-substituted oxime group and subsequent rearrangement of the remaining structure (P1_PRO and P3_PRO), while P2_PRO is generated by hydroxylation of the hydroxycyclohexenonic ring.

5.3.5 Real samples analysis

5.3.5.1 Paddy water target analysis

LC-MS/MS target analyses were performed on all the samples collected in the area of Valle Lomellina during the period from April to September 2015 (509 samples) and from April to October 2016 (558 samples), in order to identify the three herbicides and their degradation products.

The different flooding techniques adopted in the fields did not highly influence the concentration of the pesticides since they showed similar values among each other in both 2015 and 2016 campaigns. The pesticides reached their maximum concentrations one day after the administration (indicated hereafter in the plots by the solid red line) and in the day after the application of the same pesticide in the adjacent field (indicated hereafter in the plots by the dashed yellow line). This may be due to pesticides drift: the herbicides may fall directly onto the crops of the surrounding fields not only due to the speed and direction of the wind, and to the aspects related to their distribution mode (diameter of the sprayed drops, pressure of the sprayer, distance between the crop and the nozzles, type and orientation of the nozzles) but also to the direction of the water flow, which is directed from north-west towards south-east in the considered paddy area, as supposed by the geologists involved in the project. In 2015 the peak concentration of IMA was the highest (16.0 µg L^{-1}), followed by OXA (6.0 µg L^{-1}), while PRO was not detected in any sample (Fig. 5.6). In the successive year IMA concentration trends were confirmed, while no PRO and OXA were found in all the analyzed samples. At the end of both campaigns, the residual concentrations measured for the three pesticides were similar to the initial ones; in particular none of the investigated herbicides was detected at the end of the 2015 campaign (Fig. 5.6), while IMA levels remained low around 2.0 µg L^{-1} at the end of the 2016 campaign, indicating a pre-existing contamination of water due to that pollutant (Fig. 5.7). To substantiate this hypothesis, IMA levels were evaluated also in the samples

collected from the groundwater: the results indicated a constant low contamination of the groundwater, around 5.0 μg L^{-1}. Furthermore, the surface water samples collected in 2016 campaign showed higher concentration values (maximum values at 23.5 μg L^{-1} in correspondence to the IMA application day) than those observed in 2015 (maximum values at 13.3 μg L^{-1}). To explain such experimental evidences, different hypotheses were advanced. The most probable concerns the pesticide drift: farmers of the surrounding area applied the IMA to their crops and, because of the water flow direction, the pesticide could have easily reached the area under investigation, increasing its contamination.
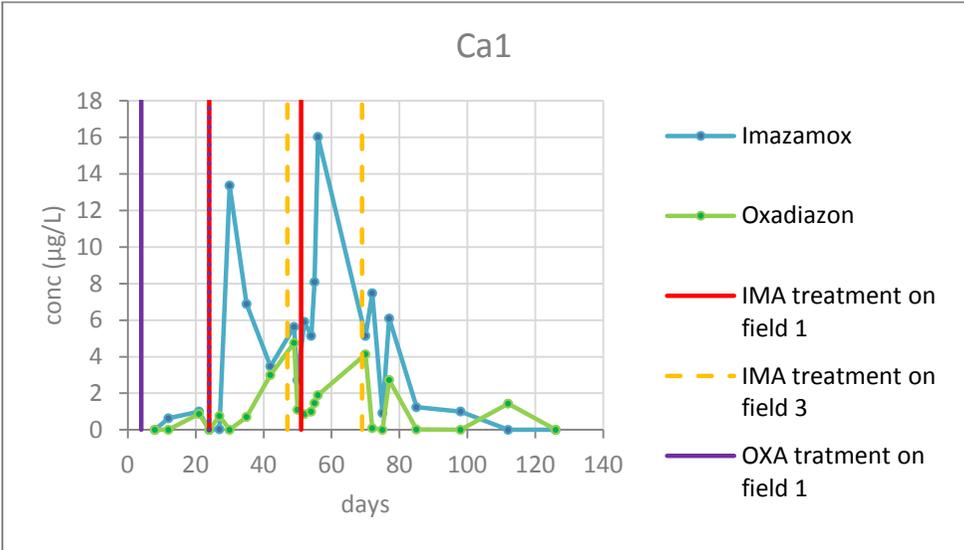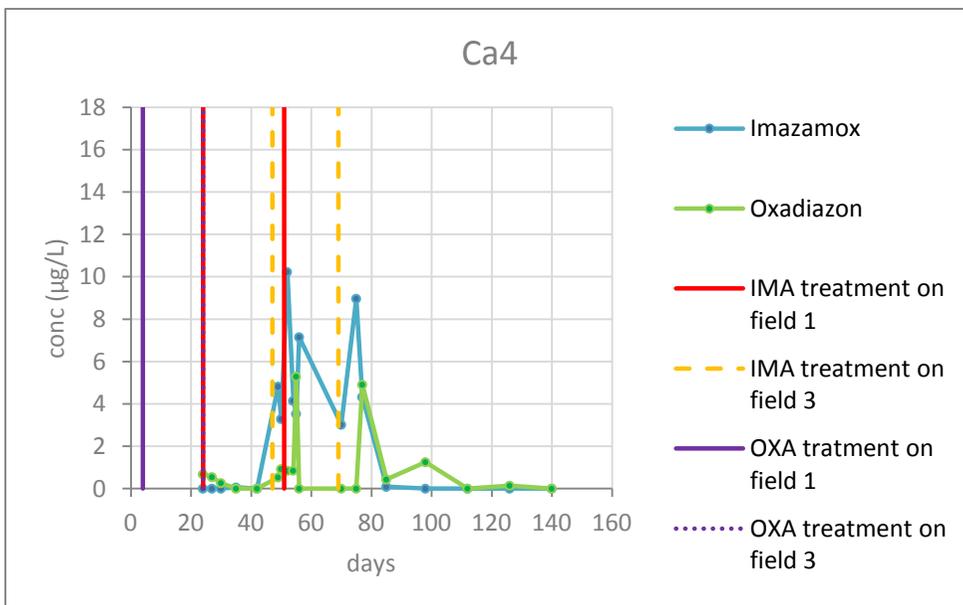

Figure 5.6: IMA and OXA 2015 concentration trends in field 1, considering the samples collected from the porous cup at 25 cm depth.
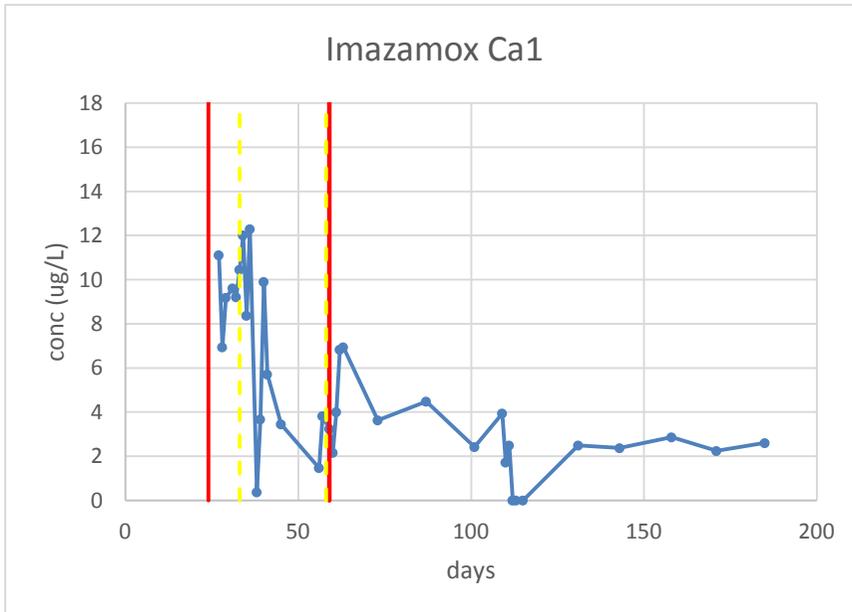
Figure 5.7: 2016 Imazamox concentration trend in field 1 (porous cup at 25 cm deep). The treatment applied on field 1 are indicated by the solid red line, while those applied to the adjacent field 3 are indicated by the dashed yellow line.

Particular attention was devoted to field 4 in which, for both campaigns, all the pesticides concentrations were always lower than in the others (Fig. 5.8). The explanation could be made by considering the particular composition and conformation of field 4 (geomorphologic data). They showed that such field is characterized by a highly permeable sandy texture and it is characterized by a slight slope. These factors may facilitate both run-off of pesticides and their possible absorption by the soil itself.

Figure 5.8: IMA and OXA 2015 concentration trends in field 4, considering the samples collected from the porous cup at 25 cm depth.

It was also considered the correlation between the herbicide and its degradation products during the campaigns. As an example, the trends of IMA and its degradation product P2_IMA in field 1 are shown, considering the Ca sampling point (Fig. 5.9 A and Fig. 5.9 B) in 2015 campaign.
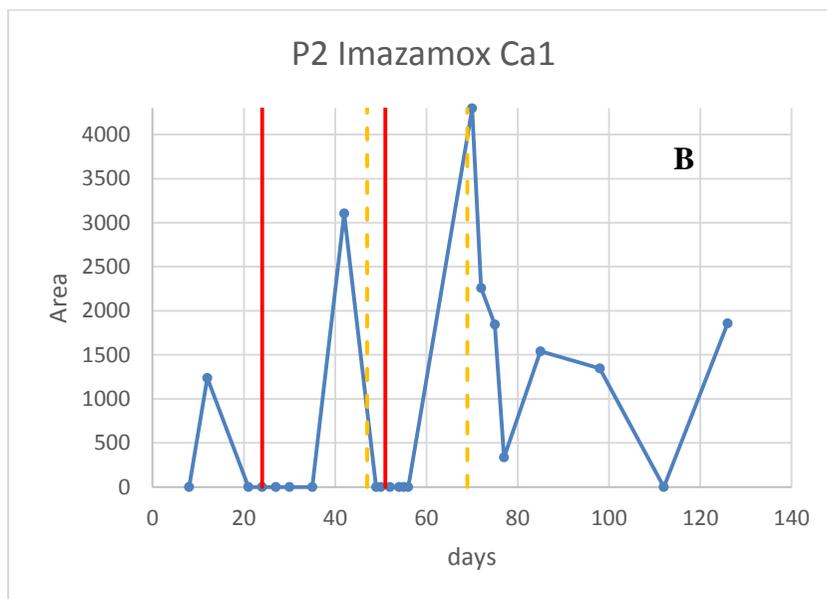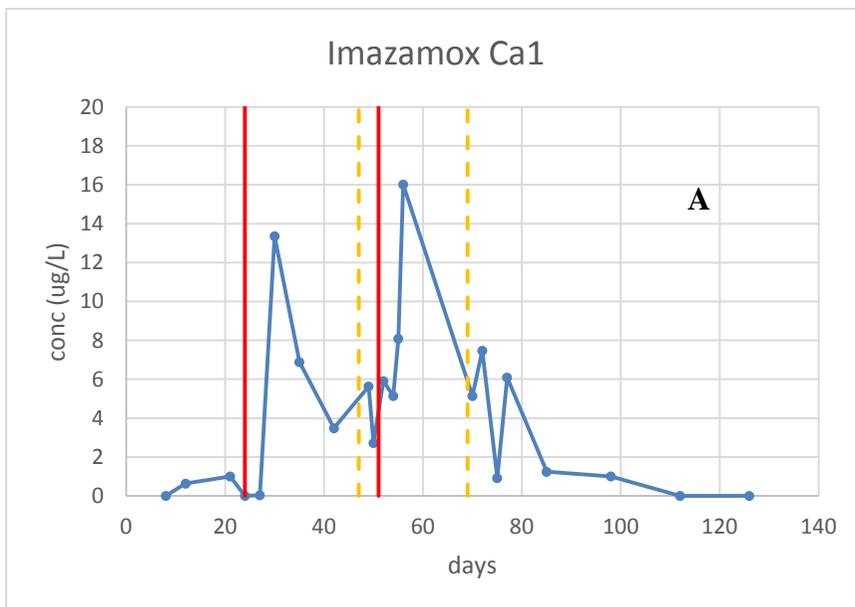
Figure 5.9: IMA concentration trend (A) and P2_IMA profile (B) in field 1, considering the sample collected from the porous cup at 25 cm of depth. The treatments applied on field 1 are indicated by the solid red line, while those applied to the adjacent field 3 are indicated by the dashed yellow line.

It can be observed how IMA concentration increases not only when field 1 is treated (indicated by the solid red line), but also when the adjacent field 3 is treated (indicated by the dashed yellow line). As time goes by from the date of the field treatment, the concentration of IMA decreases, accompanied by the formation of its degradation product P2_IMA. Such relationship is evidenced by P2_IMA trend reported in Figure 5.9 B. In fact, P2_IMA forms after IMA treatment, supporting the herbicide degradation results and the data related to its mobility in the paddy water compartment.

Furthermore, the couple of target compounds constituted by IMA and P2_IMA was adopted as a tracer to support fluid dynamic and stratification data of the fields proposed by the geologists working on the project. The results confirmed the proposed flow of water that was supposed to be coming from north-west and directed toward south-east of the considered paddy area, supporting the theory of herbicides drift due to water flows.

None of the PRO degradation products identified in paddy water solutions exposed to solar box irradiation was identified in real paddy fields. A possible explanation could be offered by the high Henry's law constant value of PRO ($1.76 \cdot 10^{-2}$ Pa$\cdot$m$^3\cdot$mol^{-1}) that suggests the rapid PRO volatilization from the paddy fields, as proposed by Tsochatzis et *al.* [27], unlike what happens during the photodegradation in the closed quartz cells of the solar box.

5.3.5.2 Paddy water non-target analysis

To offer a complete overview of paddy water composition, non-target screening was carried out on all the collected samples. The identification of new chromatographic peaks was performed once again through the help of the dedicated software MasterView™ after setting the restrictive criteria described in Paragraph 5.3.4: all the samples collected during the campaigns were compared with a blank paddy water sample collected from the same area, before fields treatments. Obviously, non-target screening is able to find out not only pesticides, but also other compounds that could be identified through LC-MS, such as other organic pollutants or characteristic compounds of natural water samples. MasterView™ software was used to explore the samples using our in-house pesticides database, consisting of the MS/MS spectra of the pesticides previously investigated. Figure 5.10 shows the workflow that allowed the identification of tricyclazole (TC) and amidosulfuron (AM) in the paddy water samples. The former is a fungicide used to prevent mycosis in the airborne organs induced by *Piricularia oryzae* [28], while the latter is a post-emergence herbicide that inhibits the acetolactate synthase activity involved in the synthesis of some amino acids necessary for weed growth. As an example, it is reported the workflow used for TC identification through library searching. The identification of a compound is based on four criteria: retention time (RT), mass accuracy, isotope pattern match and MS/MS library match. The performed direct comparison between the sample and the control, evidenced as significant the signal at *m/z* 190.0428, whose corresponding chromatographic peak is at RT=3.41 min (Fig. 5.10 A). Thanks to the information on the isotopic pattern of the detected quasi-molecular ion (Fig. 5.10 B) and accurate MS/MS spectrum (Fig. 5.10 C), it was possible to identify TC. As a mirror view, in fact, the library MS/MS spectrum of TC (grey trace, Fig. 5.10 C) is reported under the experimental MS/MS spectrum associated to the evidenced *m/z* signal (blue trace, Fig. 5.10 D), indicating the positive matching between the most abundant fragments, hence the correct attribution of the signal *m/z* at 190.0428 to

TC. Such fungicide was identified in all the collected samples, while AM was identified in the porous cup located at 25 cm deep in field 1, at the end of April. In fact, to protect rice cultivations from cleavers and other broad-leaved weed, AM has to be applied at an early age, with a height of weed of 3-20 cm.
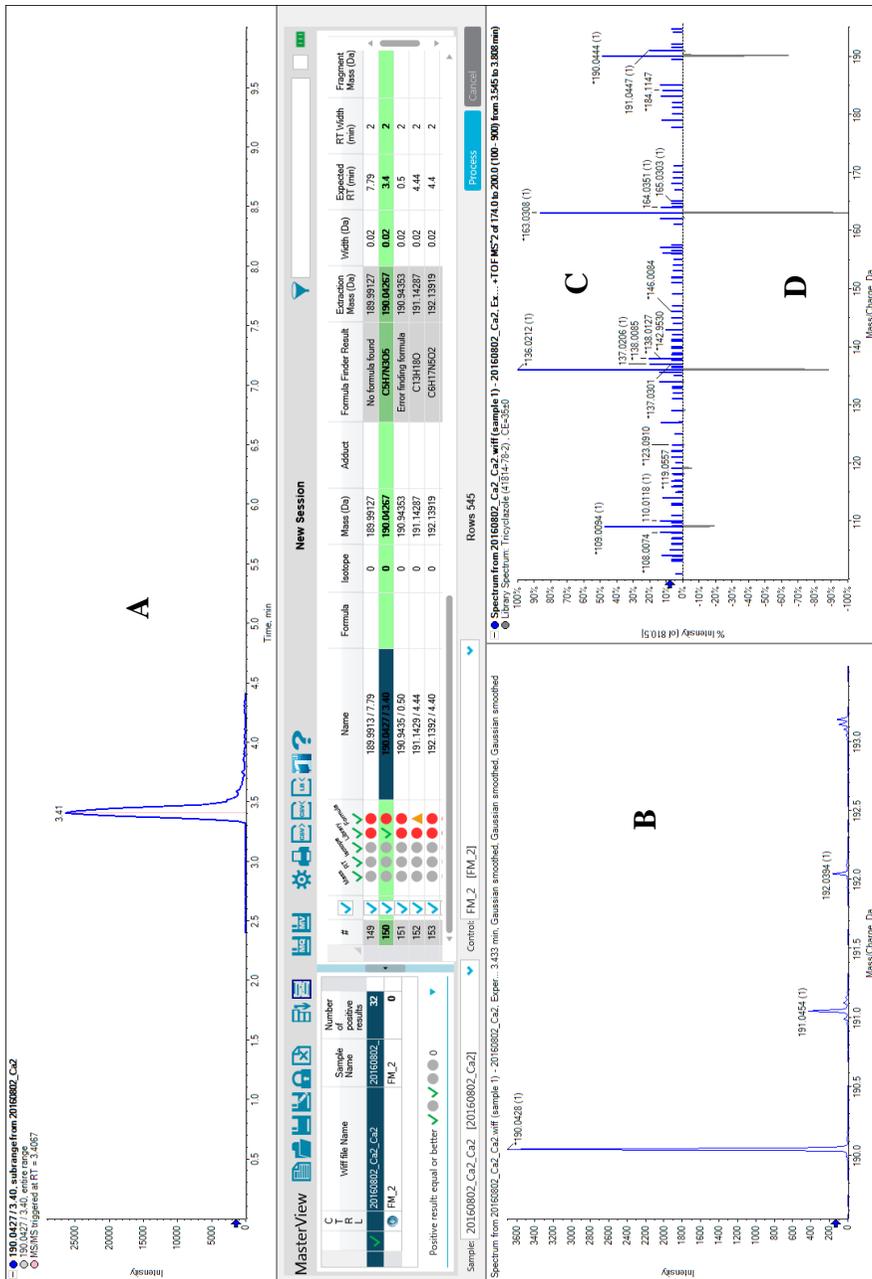
Figure 5.10: extracted ion chromatogram (A), isotopic pattern (B), accurate MS/MS spectrum (C) of the signal at m/z 190.0428 and the library matching with Tricyclazole (TC) MS/MS spectrum (D).

The presence of TC was first confirmed through the comparison with the standard, which was available in our laboratory, then it was included in the method and quantified in all the collected samples. As an example, its concentration trend is reported for paddy water samples collected from field 1 through the porous cup located at 25 cm deep, during the 2015 campaign (Fig. 5.11).

TC concentration was high at the beginning of the campaign and decreased until the end of it, showing comparable concentration levels for the considered time span.
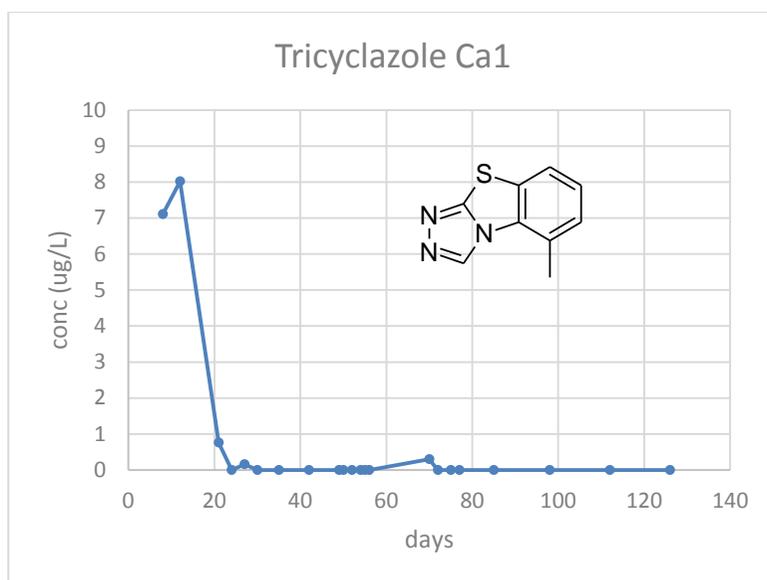
Figure 5.11: 2015 tricyclazole concentration trend in field 1 (porous cup at 25 cm deep).

Given the presence of TC in all the considered samples, it was evaluated that of its degradation products, whose structures had been elucidated in a previous project [5]. The results were negative in both the campaigns.

In addition, other two unknown species (m/z 192.1375 and m/z 230.1172), whose elemental compositions were $C_{12}H_{17}NO$ and $C_{14}H_{15}NO_2$ (labeled as U1 and U2), were identified in both years.

The two molecular formulas were then searched against ChemSpider database in order to suggest their chemical structures. Figure U1 shows, as example, the workflow used to identify U1 as N,N-diethyl-3-methylbenzenamide (DEET). U1 *m/z* signal at 192.1375 is extracted from the total ion current (Fig. 5.12 A) as it is considered as significant by the direct comparison between the sample and the control. The MS (Fig. 5.12 B) and MS/MS (Fig. 5.12 C) spectra are associated to the chromatographic peak at RT=4.44 min (Fig. 5.12 A). The information obtained is used by MasterView[TM] to compare the acquired MS/MS spectra for the extracted *m/z* signal and those present in the ChemSpider online database, used to suggest its potential molecular and structural formulas. The most probable structure proposed by the software is that of N,N-diethyl-3-methylbenzenamide (DEET), ranked first in the list suggested by the software (Fig. 5.12 D). The U1 acquired MS/MS spectra is compared with that simulated *in silico* for the same compound by the software, considering breaking and rearranging bonds rules (Fig. 5.12 F). The most abundant fragments gave positive matching with the library ones, confirming the attribution proposed by the software, hence indicating the presence of DEET in the considered sample. DEET is one of the most employed active principles in insects repellent formulations. Due to its great diffusion/commercialization, it is considered as an emerging contaminant, belonging to the personal care product (PCP) class [29, 30]. Figure 5.13 reports its chromatographic area trend in field 2, considering the sampling point of the porous cup Ca1. The maximum concentration of DEET was evidenced in the early summer period, considering the period from the middle of June the middle July.
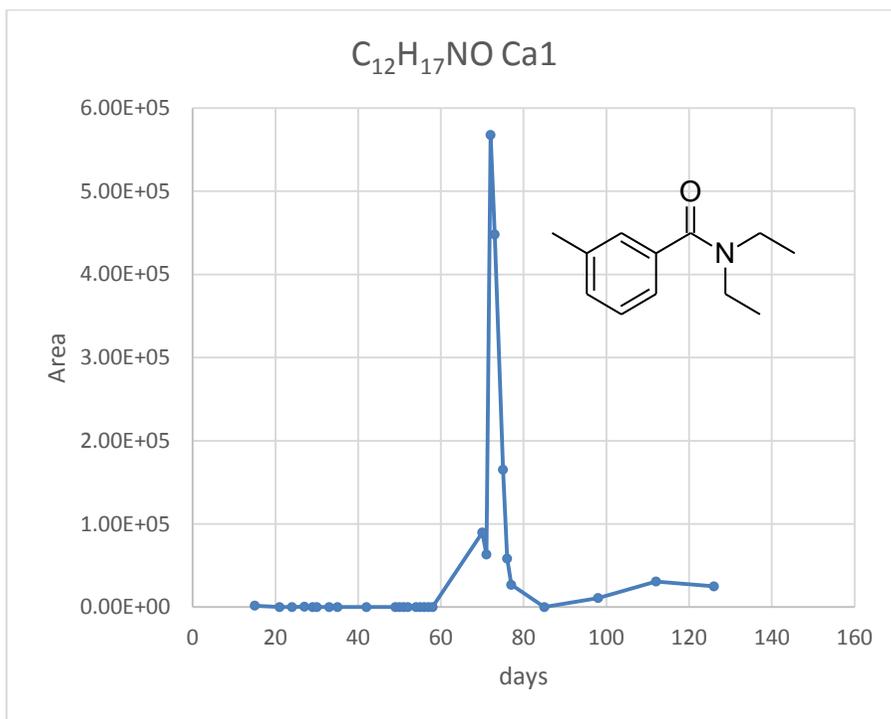
Figure 5.13: N,N-diethyl-3-methylbenzenamide (DEET) attribution: XIC of m/z 192.1375 (PI) (A), its MS (B) and MS/MS (C) spectra, the ChemSpider database suggested name (D), the proposed chemical structure (E) and the positive matching between the experimental MS/MS spectrum and that simulated *in silico* (F).
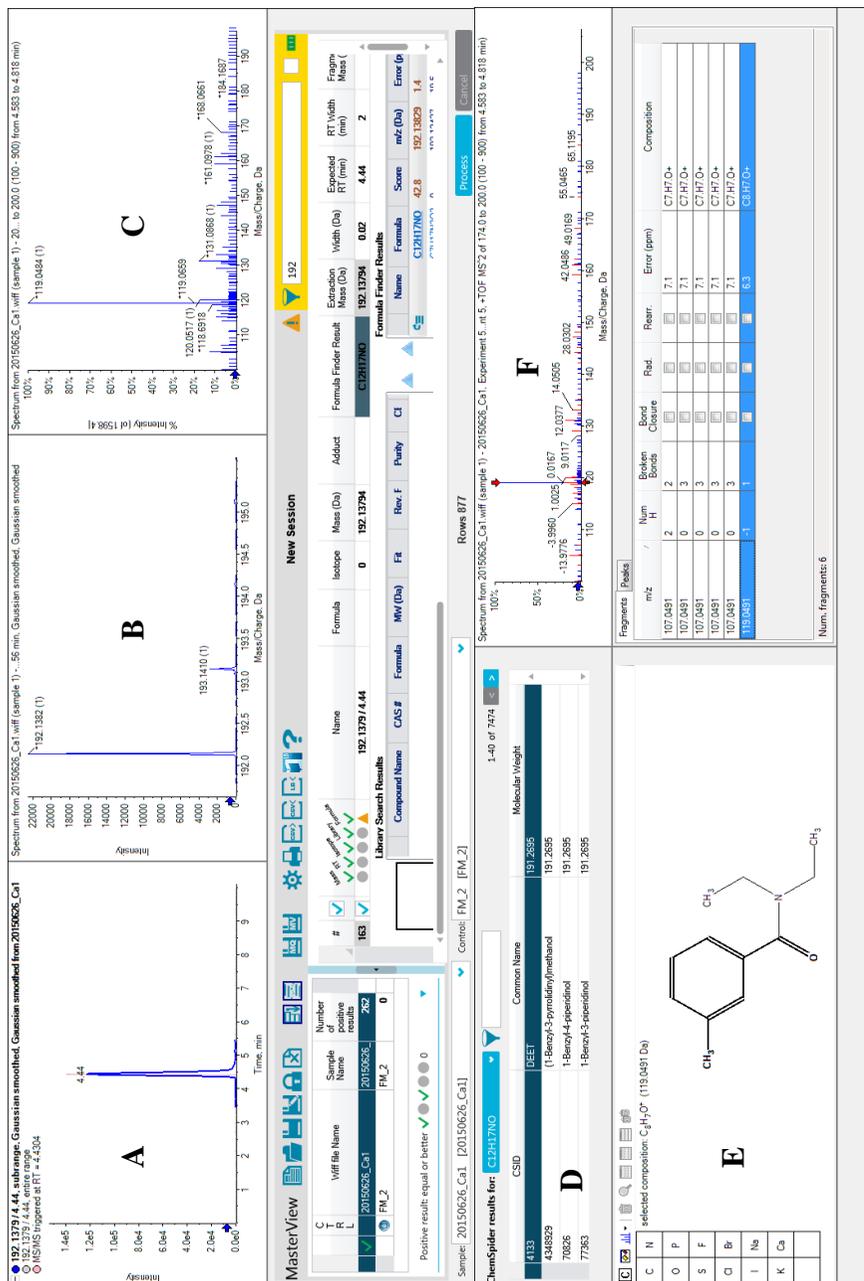
Figure 5.13: trend of DEET chromatographic area of samples collected from the porous cup Ca of the field 1.

On the contrary, U2 did not found positive matching in ChemSpider database. However, its levels were monitored during both campaigns, evidencing comparable profiles. As an example, it is reported its trend in 2015: it is interesting to notice that U2 is totally absent in the first 50 days of the campaign and then it starts to increase reaching maximum levels at day 72, keeping constant its level until the end of the 2015 study, corresponding to summer (Fig. 5.14).
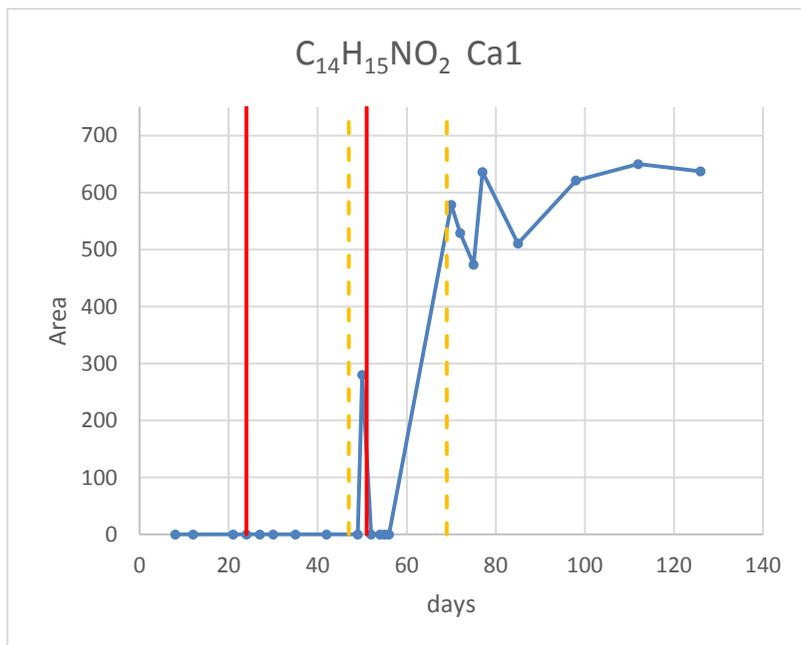
Figure 5.14: U2 ($C_{14}H_{15}NO_2$) chromatographic area trend during the 2015 campaign, considering the samples coming from the porous cup at 25 cm depth in field 1.

In addition, in 2016 campaign it was possible to identify two other unknown compounds, labeled as U3 and U4, whose molecular formula are $C_{14}H_{25}NO_4$ (*m/z* 272.1860, PI) and $C_{15}H_{25}N_5O_3$ (*m/z* 324.2024, PI), respectively. We tried to evaluate if U3 and U4 were degradation products of the pesticides employed, indicating the different treatments on the same field with different color lines: IMA is indicated by the red line, OXA by the green one and PRO by the dashed blue one.

Their structural elucidation is still in progress, but as for U2, we could follow their chromatographic area trends in all the campaigns (Fig. 5.15 and Fig. 5.16).
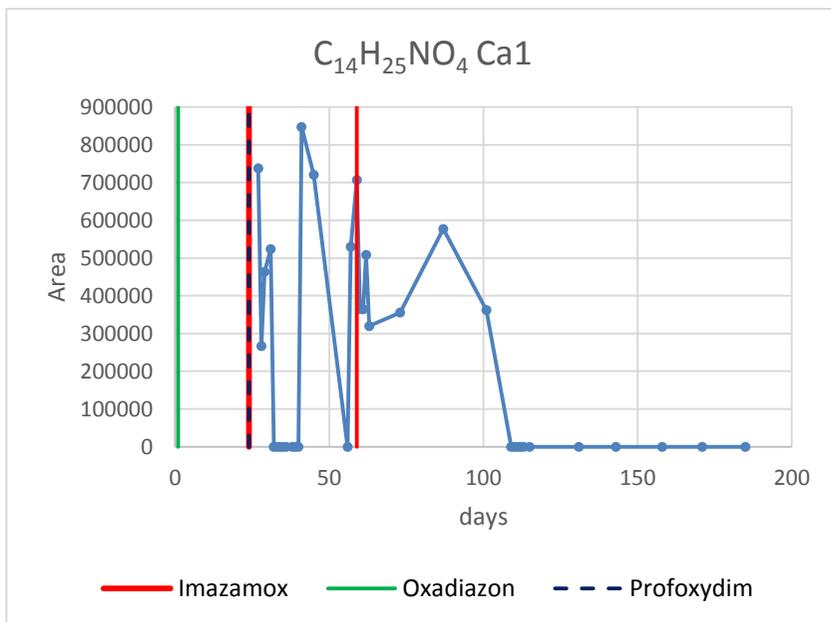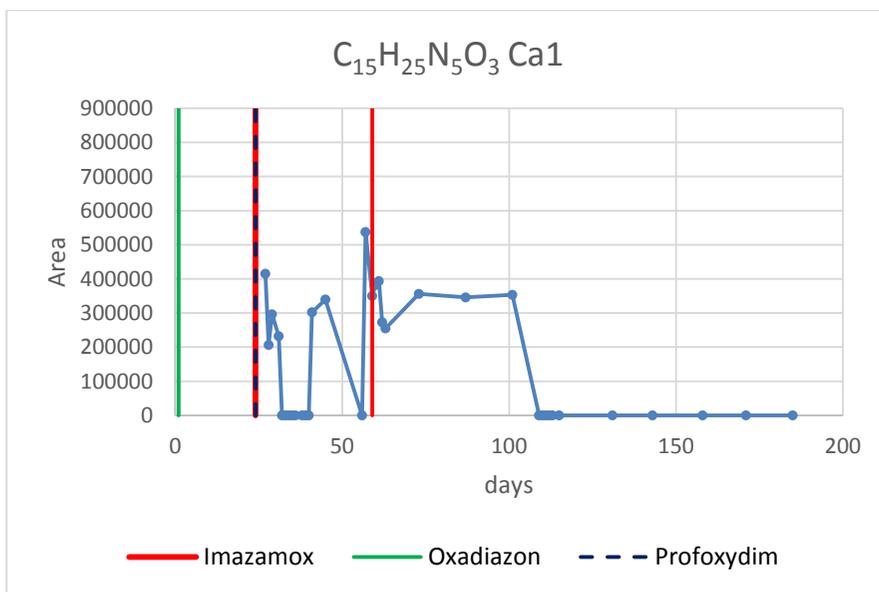
Figure 5.15: U3 ($C_{14}H_{25}NO_4$) chromatographic area trend during the 2016 campaign, considering the samples coming from the porous cup at 25 cm depth of field 1.

Figure 5.16 U4 ($C_{15}H_{25}N_5O_3$) chromatographic area trend during the 2016 campaign, considering the samples coming from the porous cup at 25 cm depth of field 1.

5.4 Conclusions

This study showed that IMA, OXA and PRO undergo to degradation in paddy water, when exposed to simulated sunlight irradiation. The three pesticides degraded in less than six hours, all following a first order kinetic reaction. Their degradation products in paddy water solutions were identified through a micro LC-MS/MS; in particular, the provided high-resolution and accurate MS data demonstrated to be crucial in the structure elucidation of the degradation products.

All the 1067 samples collected during the two campaigns were investigated through the micro LC-MS/MS developed method, using the Data Independent Acquisition mode SWATH^{TM}. The unselective SWATH^{TM} acquisition method allowed the analysis of all the components in each collected water sample, considering both target and non-target compounds, providing the complete fingerprint of all the samples and highlighting their complexity.

Non-target screening of paddy water samples characterized by thousands of peaks required the use of dedicated software able to semi-automatically screen the data in a reasonable time span. The elucidation of the chemical structures of the unknown molecules present in the water samples, in fact, was first performed with the help of an in-house pesticides library: the positive MS/MS library matching allowed the identification of TC and AM. Since the TC standard was available in our laboratory, it was quantified in all the collected samples. In particular, the presence of TC was confirmed by the data collected by local Environmental Protection Agency that identified TC in the water samples at the same levels of those evidenced by our analysis. Such evidence confirms that TC could be considered as a ubiquitous pollutant as it was already present at the beginning of the study.

Furthermore, thanks to the exhaustiveness of SWATH^{TM} analysis and the help of ChemSpider database, it was also possible to identify the emerging contaminant DEET in both campaigns in samples collected in the early summer period. Other

three species were identified in 2015 (U2) and 2016 (U3 and U4) campaigns, but their structures are still under elucidation.

In addition, some of the investigated species, such as the couple IMA/P2_IMA were used as tracers to support the hypothesis on fluid dynamics and soil composition predicted by the geologists involved in the project.

References

1. Sopeña F M C, Morillo E (2009) Cien Inv Agr 35: 27-42.

2. European Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directive 2000/60/EC and 2008/105/EC as regards prioritary substances in the field of water policy, Off J Eur Union (2013) L226: 1-17.

3. Gosetti F, Chiuminatto U, Zampieri D, Mazzucco E, Marengo E, Gennaro M C (2010) J Chromatogr A 1217: 3427-3434.

4. Gosetti F, Bottaro M, Gianotti V, Mazzucco E, Frascarolo P, Zampieri D, Olivieri C, Viarengo A, Gennaro M C (2010) Environ Pollut 158: 592-598.

5. Gosetti F, Chiuminatto U, Mazzucco E, Mastroianni R, Bolfi B, Marengo E (2015) Environ Sci Pollut Res 22: 8288-8295.

6. Ishizuka K, Hirata H, Fukunaga K (1975) Agr Biol Chem 39: 1431-1446.

7. Hirata H, Ishizuka K (1975) Agr Biol Chem 39: 1447-1454.

8. Chakraborty S K, Chowdhury A, Bhattacharyya A, Ghosh S, Pan S, Waters R, Adityachaudhury N (1995) J Agric Food Chem 43: 2964-2969.

9. Ambrosi D, Kearney P C, Macchia J A (1977) J Agric Food Chem 25: 868-872.

10. Bingham S W, Shaver R L, Guyton C L (1980) J Agric Food Chem 28: 735-740.

11. Harir M, Frommberger M, Gaspar A, Martens D, Kettrup A, El Azzouzi M, Schmitt-Kopplin Ph (2007) Anal Bioanal Chem 389: 1459-1467.

12. Bottaro M, Frascarolo P, Gosetti F, Mazzucco E, Gianotti V, Polati S, Pollici E, Piacentini L, Pavese G, Gennaro M C (2008) J Am Soc Mass Spectrom 19: 1221-1229.

13. European Food Safety Autority (EFSA) (2016) Peer review of the pesticide risk assessment of the active substance imazamox. 4: 4432-4548 doi: 10.2903/j.efsa.2016.4432.

14. European Food Safety Autority (EFSA) (2010) Conclusion on the peer review of the pesticide risk assessment of the active substance oxadiazon. 8: 1389-1481.

15. Chakraborty S K, Bhattacharya A, Chowdhuy A (1999) Pestic Sci 55: 943-948.

16. Kadokami K, Morimoto M, Haraguchi K, Koga M, Shinohara R (1991) Anal Sci 7: 247-252.

17. Kobayashi H, Ohyama K, Tomiyama N, Jimbo Y, Matano O, Goto S (1993) J Chromatogr A 643: 197-202.

18. Mattern G C, Liu C-H, Louis J B, Rosen J D (1991) J Agric Food Chem 39: 700-704.

19. https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=%E2%9C%93&search=Profoxydim

20. https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=%E2%9C%93&search=oxadiazon

21. https://comptox.epa.gov/dashboard/dsstoxdb/results?search=24239

22. Garrido E M, Lima J L F C, Delerue-Matos C, Borges M F M, Oliveira Brett A M (2001) Electroanalysis 13: 199-203.

23. Ishibashi M, Suzuki M (1988) J Chromatogr A 456: 382-391.

24. George D A (1982) J Assoc Anal Chem 65: 28-37.

25. Liao W, Joe T, Cusik W G (1991) J Off Assoc Anal Chem 74: 554-565.

26. Boyd-Boland A A, Pawliszyn J B (1995) J Chromatogr A 704: 163-172.

27. Tsochatzis E D, Tzimou-Tsitouridou R, Menkissoglu-Spiroudi U, Karpouzas D G, Katsantonis D (2013) Chemosphere 91: 1049-1057.

28. Padovani L, Capri E, Padovani C, Puglisi E, Trevisan M (2006) Chemosphere 62: 303-314.

29. Zenobio J E, Sanchez B C, Leet J K, Archuleta L C, Sepúlveda M S (2015) 120: 750-755.

30. Hallè C, Huck P M, Peldszus S (2015) J Am Water Works Assoc 107: E364-E379.

Chapter 6

Conclusions

This PhD project demonstrated the powerful versatility of the LC-MS/MS technique in performing target and non-target analyses of complex matrices. In particular, it both provides sensitive and accurate methods for the quantification of selected compounds (target analytes) and the identification of unknown compounds in the non-target approach. The complementarity of the different information provided by the two types of analysis was exploited in the three different projects presented in this PhD thesis. In particular, the high sensitivity and selectivity of LC-MS/MS in performing target analysis was evidenced, able to provide very low limits of detection and quantification, especially when the sMRM mode was adopted. The innovation of the new LC-MS/MS system consists of performing target and non-target analysis in a single chromatographic run. Full scan acquisition provides the collection of thousands of signals, not only coming from the compounds typical of the investigated sample (both target and non-target).

The Data Independent Acquisition mode was more sensitive in terms of detection and identification than Data Dependent Acquisition, since no restrictions are imposed: in this way, all the precursor ions are fragmented in MS/MS experiments. The approach generates thousands of signals that have to be interpreted to obtain the effective chemical composition of the analyzed sample: in order to attribute a chemical structure to the *m/z* signals found, database library searching (i.e. ChemSpider) is adopted as the first investigative instrument. The success in attributing the correct structure to the unknown compound is represented by the matching percentage (>70%) between the MS/MS acquired spectra and those provided by ChemSpider.

If none of the proposed structures shows a good matching, a meticulous interpretation of the accurate and high-resolution MS and MS/MS spectra has to be performed, evaluating also the accurate elemental composition, abundance of the isotopic cluster, number of ring and double bonds, LC retention time and Log*P*. Unfortunately, the software used for data processing are semi-automated and many compounds are not present in the libraries (like for example the degradation products), making the interpretation of the acquired data time-consuming and laborious.

The combination of target and non-target analysis offered a detailed picture of each of the investigated matrices and their digitalization, providing the possibility of re-interrogating the same sample, without running it again, thus avoiding the problems related to low amounts of samples and/or degradation. Furthermore, whenever the structure of a non-target compound is established and confirmed by standard comparisons, it is possible to add such molecule to the target list, than become object of routinary analysis and regulations.

The three projects presented in this thesis demonstrate how liquid chromatography coupled with high-resolution mass spectrometry is gaining importance as it represents one of the most attractive techniques to screen and identify unknowns: its routinary employment will ensure safety of cosmetic formulations, food manufacturing and will improve environmental monitoring.

List of Publications

1. F. Gosetti, **B. Bolfi**, U. Chiuminatto, M. Manfredi, E. Robotti, E. Marengo "Photodegradation of the pure and formulated alpha-cypermethrin insecticide gives different products" Environ Chem Lett. 60 (2017) 1-10. DOI: 10.1007/s10311-017-0685-4.

2. F. Gosetti, U. Chiuminatto, S. Martinotti, **B. Bolfi**, E. Ranzato, M. Manfredi, E. Marengo "Characterization of the Volatile and Nonvolatile Fractions of Heartwood Aqueous Extract from Pterocarpus marsupium and Evaluation of Its Cytotoxicity against Cancer Cell Lines" Planta Med. 82 (2016) 1295-301. DOI: 10.1055/s-0042-104659.

3. F. Gosetti, **B. Bolfi**, E. Mazzucco, M. Manfredi, E. Marengo "LC-MS/MS identification of unknown species in beverages due to interactions between dyes and other additives or due to dye degradation reactions" in "Advances in Food Analysis research", chapter 10, Haynes Ed., Nova Publishers inc, USA, ISBN: 978-1-63483-783-5.

4. F. Gosetti, **B. Bolfi**, M. Manfredi, G. Calabrese, E. Marengo "Determination of eight polyphenols and pantothenic acid in extra-virgin olive oil samples by a simple, fast, high-throughput and sensitive ultra-high performance liquid chromatography tandem mass spectrometry method". J Sep Sci, 38 (2015) 3130-3136. DOI: 10.1002/jssc.201500452.

5. F. Gosetti, **B. Bolfi**, E. Marengo "Identification of sulforhodamine B photodegradation products present in nonpermanent tattoos by micro liquid chromatography coupled with tandem high-resolution mass spectrometry". Anal. Bioanal. Chem. 407 (2015) 4649-4659. DOI 10.1007/s00216-015-8667-5.

6. F. Gosetti, U. Chiuminatto, E. Mazzucco, R. Mastroianni, **B. Bolfi**, E. Marengo "Ultra high performance liquid chromatography tandem high-resolution mass spectrometry study of tricyclazole photodegradation products in water". Environ Sci Pollut Res. 22 (2015) 8288-8295. DOI 10.1007/s11356-014-3983-4.

Acknowledgements

DECLARATION AND AUTHORISATION TO ANTIPLAGIARISM DETECTION

DECLARATION AND AUTHORISATION
TO ANTIPLAGIARISM DETECTION

The undersigned Bolfi Bianca student of Chemistry&Biology Ph.D. course (XXX Cicle)

declares:

- to be aware that the University has adopted a web-based service to detect plagiarism through a software system called "Turnitin"
- her Ph.D. thesis was submitted to Turnitin scan and reasonably it resulted an original document, which correctly cites the literature

acknowledges:

- her Ph.D. thesis can be verified by his Ph.D. tutor and/or Ph.D Coordinator in order to confirm its originality.

Date: 15/11/2017   Signature: Bianco Bolfi