UNIVERSITY OF VERONA

DEPARTMENT OF MEDICINE AND PUBLIC HEALTH

GRADUATE SCHOOL IN TRANSLATIONAL BIOMEDICAL SCIENCES

PhD COURSE IN FORENSIC SCIENCE AND MEDICINE

XXII CYCLE

PhD THESIS

DEVELOPMENT AND EVALUATION OF NEW STRATEGIES FOR THE GENERAL UNKNOWN TOXICOLOGICAL SCREENING IN BIOSAMPLES USING HIGH RESOLUTION MASS SPECTROMETRY

S.S.D. MED/43

Supervisor: Prof. Franco Tagliaro

Tutor: Prof. Aldo Polettini

PhD Student: Eloisa Liotta

16th April 2010

TABLE OF CONTENTS

1. ABSTRACT

Introduction: The screening for Pharmaco/Toxicologically Relevant Compounds (PTRC) in biosamples has benefited a lot from MS techniques. The so-called *library search* approach has enabled the development of effective identification methods based on comparison of unknown and reference spectra. However, a downside of this approach is the limited number of reference mass spectra, particularly in the case of LC-MS where in-house/commercial databases typically include not more than one thousand compounds. High resolution MS (HRMS) enables the identification of a molecular formula (MF) through the accurate measurement of mass and isotopic pattern. However, the identification of an unknown compound starting from MF requires additional tools: (a) a database associating MFs to compound names, and (b) a way to discriminate between isomers.

Aims: To evaluate the ability of combined novel "metabolomic"/chemometric approach to reduce the list of candidate isomers.

Methods: Urine/blood/hair samples collected from real positive cases were submitted to a screening procedure using ESI-MSTOF (positive ion mode) combined with either capillary electrophoresis or reversed phase LC (RPLC). Detected peaks were searched against a PTRC database (ca. 50.500 compounds and phase I and II metabolites) consisting of a subset of PubChem Compounds. In order to discriminate between compounds with identical MF first a filter based on a "metabolomic" approach was applied. Starting from the mass of the unknown compound, defects/increments corresponding to pre-defined biotransformations (*e.g.* demethylation, hydroxylation, glucuronidation, etc.) were calculated and the corresponding mass chromatograms were extracted from the total ion current (TIC) in

order to search for metabolite peaks. For each candidate in the retrieved list, the number of different functional groups in the molecule (N,O,S-methyls, hydroxyls, acetyls, etc.) was automatedly calculated using E-Dragon software (Talete srl, Milan, Italy). Then, the presence of metabolites in the TIC was matched with functional groups data in order to exclude candidates whose structure was not compatible with observed biotransformations (*e.g.* loss of methyl from a structure not bearing methyls, glucuronidation on a structure not bearing any site susceptible to conjugation). A further filter was then applied based on a mathematic model correlating RPLC relative retention time (ISTD nalorphine) with a number of parameters estimated for each candidate compound starting from the Simplified Molecular Input Line Entry Specification (SMILES), including the predicted octanol/water partition coefficient (Log*P*).

Results: The procedure was tested on 121 compounds detected in real positive samples, including drugs of abuse (*e.g.* cocaine, opiates, MDMA), anticonvulsants (*e.g.* gabapentin, carbamazepine), benzodiazepines (*e.g.* flurazepam), antidepressants (*e.g.* citalopram, trazodone, fluoxetine, amitriptyline, venlafaxine), phenothiazines (*e.g.* chlorpromazine, promazine, pericyazine), antipsychotics (*e.g.* amisulpride), antihistamines (*e.g.* cetirizine), beta-blocker (*e.g.* bisoprolol), anti-retroviral agents (*e.g.* emtricitabine, tenofovir), acetyl-cholinesterase inhibitors (*e.g.* rivastigmine), histamine H_2-receptor antagonists (*e.g.* ranitidine), and their phase I metabolites. Overall, the mean list length (MLL) of compounds was 6.71 ± 4.66 (median 6, range 1-28) before the application of the metabolomic approach and was shortened to 3.94 ± 3.07 (median 3, range 1-17) after. For RPLC-HRMS data the MLL was shorted from 6.02 ± 3.49 (median 6, range 2-21) to 3.42 ± 3.03 (median 3, range 1-17) after the metabolomic filter and to 3.09 ± 2.03 (median 2, range 1-9) after

the chemometric approach. The application of both filters allowed a reduction of the MLL to 2.14 ± 1.63 (median 2, range 1-9).

Conclusion: HRMS allows a much broader search for PTRC than other screening approaches. The combined “metabolomic”/chemometric approach significantly reduces the list of candidate isomers.

2. INTRODUCTION

General unknown screening (GUS) for drugs and toxic compounds, or "systematic toxicological analysis (STA)" or "comprehensive screening of drugs and toxic compounds", embraces all the analytical methods, or their combinations, developed with the objective to detect and identify unknown xenobiotics (*i.e.* compounds or substances extraneous to the body: therapeutic drugs, drugs of abuse and toxic compounds, including their metabolites) in biological matrices (Lambert W.E. et al., 1995; Drummer O.H., 1999; Marquet P., 2006).

The unequivocal identification of unknown pharmaco/toxicologically relevant compounds (PTRC) in biosamples represents an hard challenge for analytical toxicologists. The problem is substantially related both to the exceedingly broad range of PTRC, considering that there are several differences of polarity, molecular weight (MW), chemical/thermal stability and pKa, and to the high number and amount of interferences in the substrate as well. In fact, drugs in the range of parts-per-million and parts-per-billion should be identified even if excessive amounts of impurities exhibit similar structure and features (Stimpfl T. et al., 2000). As a consequence, high specificity is demanded beside sensitivity, because many other xenobiotics or endogenous biomolecules may interfere with the detection of the unknown analyte.

STA has a number of relevant applications, both in clinical and forensic settings. In the field of clinical toxicology, either the diagnosis or the exclusion of an acute or chronic intoxication is of great relevance. Furthermore, patients addicted to alcohol, medicaments or illegal drugs have to be monitored. For the determination of clinical brain death as a prerequisite for explantation of organs, the presence of drugs that

may depress the central nervous system must be excluded. Similar problems arise in forensic toxicology, where the proof of drug abuse or of poisoning are important tasks. Furthermore, drugs that may alter behaviour, thus impacting either on penal responsibility of a defendant or impairing the ability to carry out specific tasks, *e.g.* driving (driving under influence diagnosis) or performing specific duties at work (Workplace Drug Testing) must be searched for in body fluids or tissues. Performance-enhancing drugs must be screened for in doping control. In fact, hundreds of chemically and pharmacologically different compounds are currently included in the list of prohibited substances of the World Anti-Doping Agency (WADA) (Strano-Rossi S. et al., 2005; Kolmonen M. et al., 2007).

STA has also important applications into the screening for xenobiotics in foodstuffs/environmental matrixes, as drugs and pesticides residues may impact on human health and have to be monitored by law before agricultural product are put on the market. Over 600 pesticide active ingredients are currently in legal use in Europe (Ferrer I. et al., 2007; Thurman E.M. et al., 2006), not to speak of the huge number of drugs used in zootechnics. Even in cases where the drug(s) involved is/are known or strongly suspected, STA must be carried out in order to search for other xenobiotics that may have contributed to the observed biological effect (*e.g.* impairment, intoxication, death).

In most of these applications, currently a multi-target analysis is usually carried out instead of a GUS (Hopfgartner G. et al., 2002). This is due both to the limited availability of pure standards of xenobiotics and/or of their reference data (retention/spectral data) and to the need to reach the required limits of detection. Therefore, the availability to perform a broader screening with no *a priori* selection of

detectable compounds that may be detected and identified would allow a significant improvement in analytical toxicology.

The aim of my work was to investigate novel strategies for general unknown screening applications based on the accurate measurement of mass and isotopic pattern.

3. STATE OF THE ART

Whenever in analytical chemistry unknown compounds must be screened and reliably identified (prior to quantification), particularly in small amounts and/or in complex matrices, mass spectrometry hyphenated with separation techniques are required (Maurer H.H., 2006; Drummer O.H., 2007) for general unknown screening procedures the full-scan modality is obviously the method of choice. The simpler search strategy includes the extraction of mass chromatograms of ions specific for a compound or a drug class in order to detect their presence (*i.e.* chromatographic peak), followed by the search of the full scan mass spectrum corresponding to the peak apex with or without background subtraction.

In the past, PTRC screening was widely marked on GC/MS analysis (Maurer H.H. et al., 2000; Soriano T. et al., 2001). In fact, the combination of the high separation power of GC with the high detection power and selectivity of MS enabled the overcoming of the intricacy of GC (*e.g.* the indispensable analytes extraction from the aqueous matrices and need of polar compounds derivatization) (Rivier L., 2003; Polettini A. et al., 1998). Furthermore, both the availability of an standardized ionization technique (electron ionization [EI]) and the natural simplicity of the information in a mass spectrum, enabled the progressive development of large databases collecting a wide range of reference mass spectra (Schauer N. et al., 2005; http://www.nist.gov, 2007; http//eu.wiley.com, 2007), some of them specifically dedicated to PTRC.

At the start of the 1990s, the introduction of atmospheric pressure interfaces (API) provided an important enhancement in the robustness of the liquid chromatography (LC) to MS hyphenation, allowing a new impulse to research in toxicants screening

(Maurer H.H., 2007; Weinmann W. et al., 2000). The great versatility of LC, due to its capability into identify also high molecular weight, polar and thermolabile compounds, contributed to the extension of the identifiable compounds range (Bogusz M.J., 1999; Ferrer I. et al., 2009). In the field of PTRC screening, several groups of research investigated the library search approach based on LC/MS apparatus by using spectral libraries specifically devoted to PTRC (Mueller C.A., 2005; Dresen S. et al., 2010). Nevertheless, in contrast to GC/MS where ionization and fragmentation processes can be very steady when standardized, LC/MS techniques show a high degree of complexity owing to the different parameters that have to be optimised (Gergov M., 2004). Consequently, the development of appropriate spectral libraries for LC/MS met several obstacles. First, soft ionization that characterizes API techniques determines the poor fragmentation of the pseudomolecular ion resulting in a loss of structural information. However, the enhancement of the fragmentation reaction may be obtained by the application of tandem mass spectrometry (MS/MS) techniques or even through the so-called in-source collisionally induced dissociation (CID) (Thurman E.M. et al., 2007). Nevertheless, limitations affect both these approaches. In the first fragmentation strategy, the precursor ions subjected to CID must be a priori defined, thus implying a selection and a limitation in the range of identifiable compounds; while, in the second the chemical noise may heavily contaminate the mass spectra. However, the limitation of MS/MS screening previously described may be overcome by the recent implementation of data-dependent acquisition routines able to trigger MS/MS fragmentation of all precursor ions that reach a given abundance threshold.

Unfortunately, the “library search” strategy must tackle another relevant problem: the limited reproducibility at the interlaboratory level of product ion spectra and of API full

scan spectra as well (Bogusz M.J. et al., 1999). In order to overcome this further obstacle, tuning procedures have been proposed. As a matter of fact, the difficulty of standardizing the ionization/fragmentation processes has much limited the progress of spectral databases for LC/MS systems, differently from what happened for GC/MS techniques. In fact, libraries of reference EI mass spectra containing up to more than two hundred thousands spectra are available for GC/MS applications (Aebi B. et al., 2002), whereas LC-MS reference spectral databases rarely exceed one thousand compounds. Some additional pitfalls characterizing in-house libraries should be also considered:

- normally, a library contains only few specific ions, and not the full mass spectrum;
- big mistakes are sometimes unfortunately included in the reference spectrum;
- a low match may occur cause of the differences in the ions signal intensity due to the instrumental diversity;
- a mismatch may happen for compounds co-eluting with interfering substances;
- the total number and type of compounds that are present in a reference library, is substantially predetermined (selection bias) (Rivier L., 2006).

Recently, time-of-flight (TOF) mass spectrometers have become available for hyphenation with GC and LC, opening up new perspectives in PTRC screening. Furthermore, also capillary electrophoresis (CE), with its good resolution and rapid method development, has become an alternative separation technique for forensic laboratories, representing a possible valid solution for the coupling with TOF instrumentation (Gottardo R. et al., 2007; Gottardo R. et al., 2008).

TOF analysers enable the sensitive acquisition at high mass resolution and the measurement of mass at high performance accuracies of few ppm. This peculiarity,

combined with the accuracy in the determination of the isotopic pattern (IP), provides the unequivocal identification of the chemical composition of small ions with a mass to charge ratio (m/z) in the range 200-500. In the field of clinical/forensic toxicology, several research groups have demonstrated the adequacy of LC/TOF systems for screening purposes (Thurman E.M. et al., 2004; Ojanpera I. et al., 2005; Ferrer I. et al., 2005; Martnez-Bueno M.J. et al., 2009; Peters R.J.B. et al., 2010). The main advantage of this approach is the hard fact into avoid building-up mass spectral libraries. However, the other side of the coin is that this identification strategy does not discriminate between isobars, *i.e.* compounds with the same chemical formula (CF).

Therefore, screening procedures based on high-resolution mass spectrometry necessitate additional tools:

- a PTRC database consisting of a compounds list names/associated molecular formulas. In fact, high-resolution MS software allows the search for CFs through the determination of the match between the calculated mass/IP and the mass/IP measured for the unknown, but does not provide any information on the structure and identity of the compound.
- a way to discriminate between compounds with identical CF.

As regards the first problem, in the recent years, several research groups have been working in the attempt to create suitable databases (Hayashida et al., 2009).

Recently, both in-house (Gergov M. et al., 2001) or commercially available databases have been used for general screening purposes. Particularly, a database of 735 PTRC including therapeutic and illicit drugs, designer drugs and metabolites was developed by Ojanpera and colleagues (Ojanpera S. et al., 2006). A further improvement to research was given by the creation of another database consisting in

124 doping agents (70 metabolites) by the same group (Kolmenen M. et al., 2007). As previously mentioned, commercial databases containing CFs of PTRC are also available:

- The Merck Index contains over 10.000 different compounds but no metabolites are present (O'Neil M. et al., 2006). This represents a serious failure for analytical toxicology cause no information could be obtained for compounds undergoing extensive biotransformation.
- The MPW library includes over than 7.800 PTRC and also metabolites (Maurer H. et al., 2007). Unfortunately, this database is primarily GC-oriented and, therefore, of limited usefulness for screening approaches based on either LC or CE-TOF MS technology. At present, it is well known that the high versatility of LC or CE as separation techniques for MS coupling are better suited for general screening respect to GC (Polettini A. et al., 2006; Gottardo R. et al., 2007; Patel B.N. et al., 2008).
- The NISTChemistry Webbook database schedules about 70.000 compounds (http://www.nist.gov, 2007). It is freely consultable through the web and includes several metabolites (although not glucuronides), but also non-PTRC compounds (*e.g.* petroleum derivatives, cosmetics,...) are comprised.
- A further available opportunity is the PubChem Compound, a component of the U.S National Institutes of Health's Molecular Libraries Roadmap Initiative (www.pubchem.ncbi.nlm.nih.gov, 2007). This consists in an extensive database of biologically active small molecules containing over than 10.000.000 records collected from different sources (*e.g.* EPA, Cambridge Soft, Discovery Gate, ChemIDPlus,...) (Hill D.W., 2009). This particular database is too large to be utilized for an effective general unknown

screening. However, it gives the possibility to create subset "*ad hoc*": For example, 270.000 substances are listed under the category "toxicology". PubChem is of freely access through the web and also freely downloadable as a text file. This feature offers the opportunity to create appropriate and specific standalone searchable databases, which are extremely useful for procedures based on HRMS general unknown search.

As regards the second issue, *i.e.* the discrimination between isobars, apart from some attempts of predicting the fragmentation on the basis of the structure of the candidate compound, which so far have proved to be not applicable in the routine, the problem currently remains unsolved.

4. THEORETICAL PRINCIPLES

4.1. SEPARATION TECHNIQUES

Selective chromatography or high efficient electrophoresis represent feasible separation techniques for coupling with mass-selective detection.

These hyphenations provide the versatility needed for several applications, from the target analysis of known molecules to the identification of the unknown compounds in the general unknown screening.

4.1.1. CHROMATOGRAPHIC SEPARATION

Chromatography is a separation process based on the distribution of analytes between two phases: the stationary phase and the mobile phase. The analytes strongly interacting with the stationary phase are retained longer in the system with respect to analytes weakly interacting with the stationary phase.

The parameter describing the behavior of a compound in a chromatographic system is the equilibrium constant K_D *i.e.* the ratio between the concentrations of the analyte in the stationary and in the mobile phases.

In order to achieve the chromatographic separation of the compounds of interest, the partition coefficient should be different for each analyte. All the chromatographic systems are constituted of a stationary and mobile phases adequately chosen in order to improve the difference in the retention behavior of the analytes, *i.e.* to enhance the separation power of the system.

The separation outcome is primary influenced by:

- ✓ the basic mechanisms defining the chromatographic process (*e.g.* affinity, ion exchange, weak interactions, polarity,…);
- ✓ more general phenomena acting against the analytes separation and causing peak tailing (*e.g.* diffusion).

A more practical way to describe the retention behavior of a given analyte in a given chromatographic system is the absolute retention time (RT), that is the time required by the analyte to be eluted from the system. This is actually made of 2 components, the retention time of an unretained compound (T_M) and the time spent by the analyte in the stationary phase (T_R). In order to reduce variations of RT as a consequence of even slight modifications in the separation conditions (*e.g.* temperature or pressure differences), the retention time relative to a given internal standard (RRT) is preferred, where: RRT= RT analyte/ RT internal standard.

Chromatography is a very versatile technique enabling the separation of gases and volatile compounds by GC, and non-volatile, polar, and high molecular weight substances by LC (Gaillard Y., 1997; Williams T.A., 1999). The latter has quite interesting features for GUS purposes as it enables the direct injection of diluted aqueous substrates (*e.g.* urine) and the simultaneous separation of the whole metabolic profile of a drug, from the parent compound up to the polar conjugated metabolites (*e.g.* glucuronides) with no need for hydrolysis and derivatization procedures.

The setup of a high pressure liquid chromatographic (HPLC) gradient system is described in Figure 1 and consists of:

- the tanks containing the different components of the mobile phase;
- a mobile phase mixing/pumping system;
- a system for sample injection (autosampler);

- a separation column maintained in a thermo-regulated compartment;
- the detector (UV lamp, mass spectrometer);
- a computer system for controlling acquisition parameters and processing data.

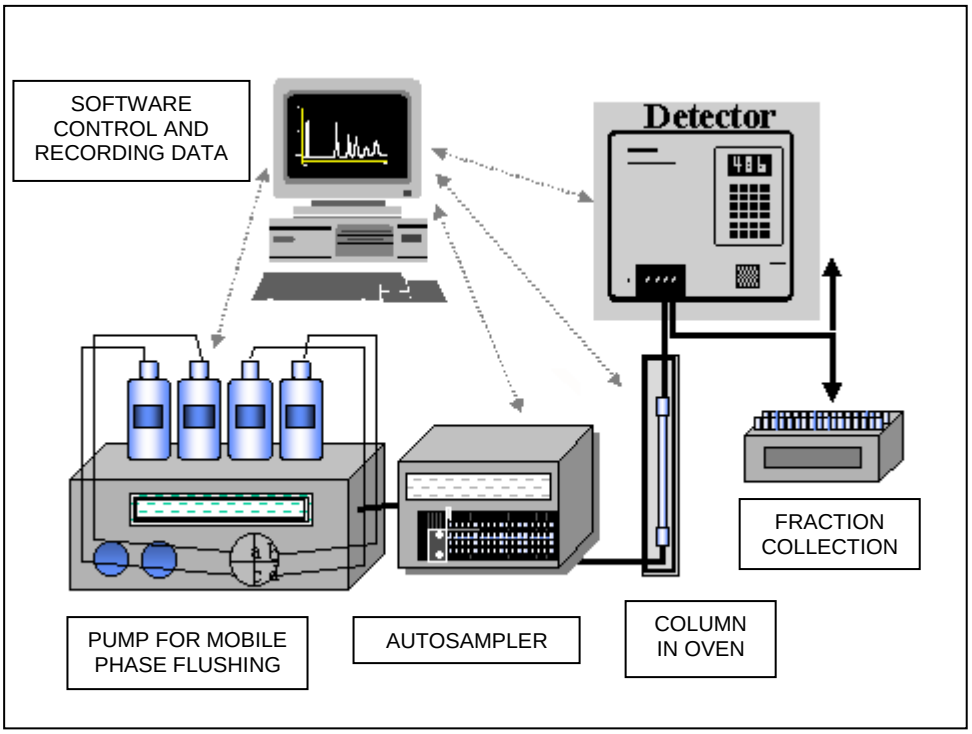


Figure 1. *Principal components of a liquid chromatographic system.*

HPLC columns are typically stainless steel pipes, with a variable internal diameter (1-4 mm or less), with lengths usually in the range 15 - 50 cm and holding pressures until 5.5 x 10^7 Pa. Their stationary phase usually covers small particles (3-5 µm diameter) packed inside the column. Several types of **stationary phases** are available. In reverse phase liquid chromatography (RPLC) an apolar stationary phase is associated to a polar mobile phase (mixtures of acetonitrile and methanol with water are commonly used). RPLC is the most common chromatographic approach as it is suitable for polar substances such as biomolecules and drug metabolites.

4.1.2. LC/MS HYPHENATION

Several studies carried out by LC/MS instrumentations demonstrated the suitability of this hyphenation for clinical and forensic toxicology applications (Maurer H.H., 2004; Lacorte S. et al., 2006; Martnez-Bueno M.J. et al. 2007).

In LC/MS system volatile buffers or additives (*e.g.* ammonia, acetic acid, formic acid, ammonium formate or ammonium acetate) are required, in order to avoid the solvents precipitation in the ion source causing reduction of the signal intensity and interface clogging. In addition, the size of the column, the flow rate and buffer elements should be adapted to the ion source. In particular, narrow-bore columns and, consequently, low flow rates are recommended for ESI (internal diameter in the range of 1-2 mm), whereas wide-bore columns using rather high flow rates of the mobile phase are preferable with APCI interfaces. In fact, ESI efficiency depends on the concentration more than on the mass flow of analyte per unit time, *i.e.* the higher is the concentration of the analyte in the mobile phase and the higher will be the analyte signal. On the contrary, APCI process depends more on a mass ionization process, and a higher amount of analyte introduced will result in an increased signal intensity. Therefore, In ESI applications, low flow rates (around 0.2 ml/min) are preferred which, besides being suitable for high ionization efficiency, account for solvents saving and less frequent servicing of the apparatus. Nevertheless, the increase in backpressure over the column (owing to the narrow I.D.) must be handled by the pumping system (Kronstrand R. et al., 2006).

4.1.3. ELECTROPHORETIC SEPARATION

Electrophoresis is a separation technique based on differential migration of charged solutes in an electric field. Particularly, in capillary electrophoresis, analytes are

separated in small-diameter capillaries (20-100 μm) under the influence of a high electric field (10-30 kV), providing fast and efficient separation of charged compounds. Briefly, CE instrumentation consists of a capillary, a power supply and a detection system (Figure 2). Notwithstanding the "basic" instrumentation, CE can be applied in various modes of operation with different separation mechanism, defining CE as a very feasible technique. The principle describing the process of electrophoresis is the **electrophoretic mobility** equation which is given by:

$$\mu_e = q / \Pi \eta r$$

where: q = ion charge

η = solution viscosity

r = ion radius

From this equation it is evident that small and highly charged species have high mobilities, whereas large and minimally charged species have low mobilities.

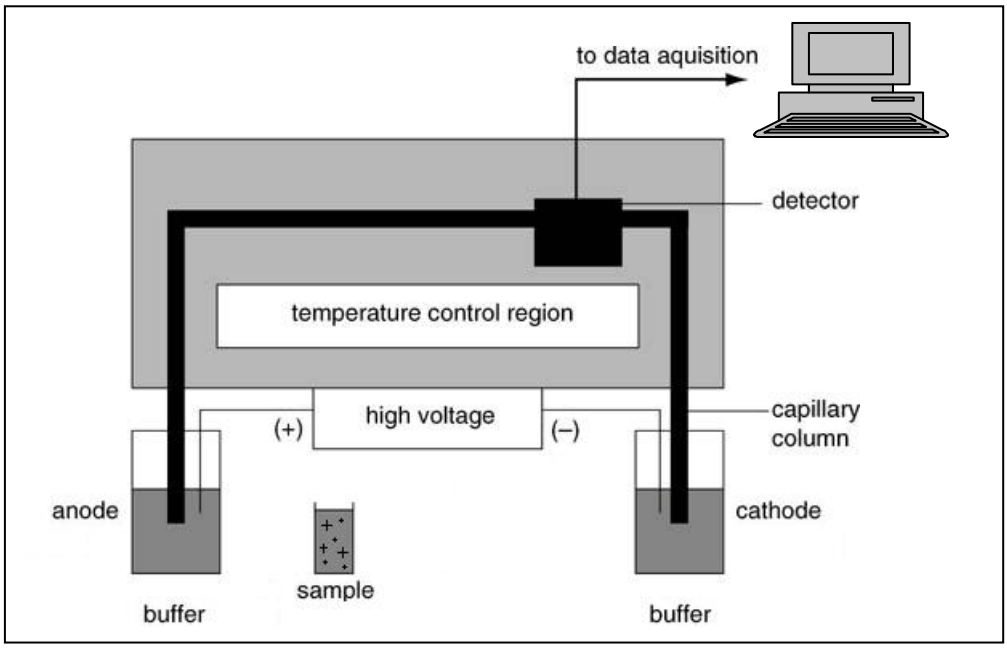


Figure 2. *Scheme of a CE system.*

The electrophoretic mobility usually found in standard tables is a physical constant, determined at the point of full solute charge and extrapolated to infinite dilution. This generally differs from that determined experimentally. Substantially, the particle mobility strongly depends on the charge to radius ratio. In this way, through the application of a potential difference, the separation of differentially charged molecules occurs, but also same charge molecules are separated due to the disparate molecular dimension. Factors affecting the electrophoretic separation are below listed and briefly commented.

1. The charge to radius ratio (typically for inorganic ions) or the charge to mass ratio (usually for macromolecules).
2. The intensity of the electric field applied: substantially represents the electrostatic force influencing the ions behavior.
3. The viscosity of the medium: influenced by the buffer composition, temperature and Joule effect.
4. The buffer composition: in order to obtain an efficient separation of the ions, the buffer solution should not chemically interact with the molecules of interest, should preserve the electric conductivity of the system and also create the adequate pH conditions.
5. The temperature: controlled temperatures are required to avoid solvent evaporation and sample degradation that may occur at high temperature.

Furthermore, a fundamental constituent of CE operation is the **electroendosmotic flow** (EOF). EOF is the bulk flow of liquid inside the capillary and is a consequence of the surface charge on the interior capillary wall. Under aqueous conditions, for a pH value of the buffer superior to 3, the EOF formation is provided. In fact, at this state, the silanol groups on the internal surface of the capillary exist in anionic form

conferring a negative charge on the capillary wall. Afterwards, the capillary filled by the buffer presents an internal surface negatively charged attracting the ions of the buffer having a positive charge. As a consequence a double electric layer is created and close to the capillary wall a potential difference achieved (zeta potential). Subsequently the application of the potential, the free cations of the layer migrate to the cathode dragging all the chemical species in the solution (positive, negative and neutral molecules) (Figure 3 A). EOF is strongly influenced by the pH of the medium: high pH values cause higher ionization of the silanol groups and consequently a greater EOF. Furthermore, the zeta potential, substantially due to the surface charge on the capillary wall, decrease at high ionic strength of the buffer and a subsequent reduction of the EOF is produced (Heiger D., 2000). However, an increase of the EOF is achieved by application of an higher potential: the Joule effect provides an enhancement of the internal capillary temperature causing the reduction of the buffer viscosity resulting in the EOF augmentation. A special feature of the EOF consists in the flat profile of its flow in the capillary (Figure 3 B). This means an important advantage since it does not directly concurs to the dispersion of the solute and does not affect the peak shape.

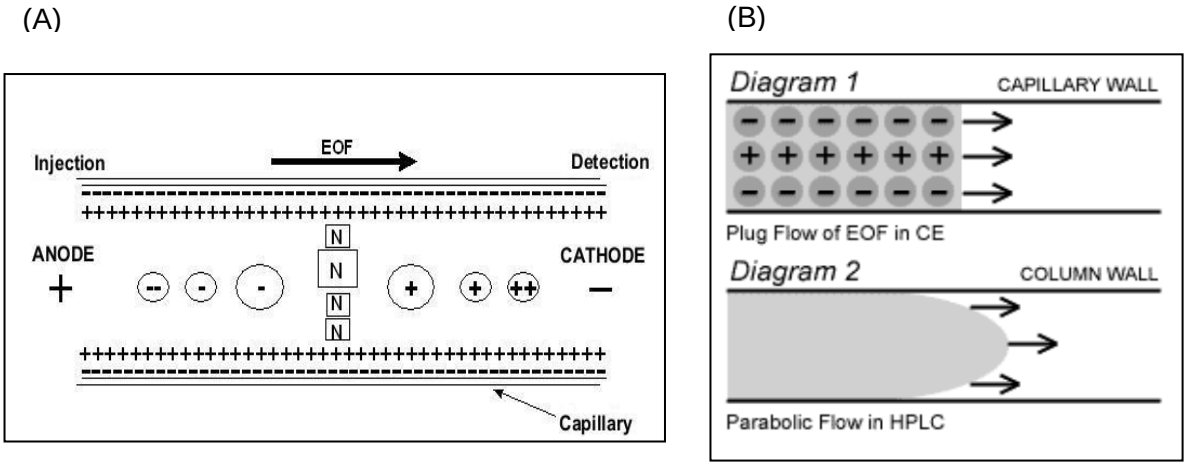


Figure 3. *(A) Negative, neutral and positive ions migration through the EOF. (B) Flow profile in a CE capillary and in a HPLC column.*

Capillary electrophoresis allows the separation of a wide range of compounds from small ions to large biopolimers (DNA, proteins,…) (Herrero M. et al., 2010).

Although CE nowadays is an efficient and versatile separation technique (Watzig H., 2003), suited for the determination of a wide variety of compounds in various matrices, the limited detection sensitivity is still a problem. This is due to the fact that the small internal diameter of the capillary offers, on one hand, low mass detection limits, but on the other hand due to the minimal volume of sample injected causes moderate concentration sensitivity. Most often to avoid band broadening, on column detection is applied. So far, the most frequently applied detection techniques are UV absorbance at single wavelength or with diode-array (DA), fluorescence, laser-induced fluorescence (LIF), electrochemical, conductivity detection and mass spectrometry (MS).

4.1.4. CE/MS HYPHENATION

CE/MS coupling was first described in the late 1980s by the group of Smith. The hyphenation allows to combine the high separation efficiency of the CE with the high sensitivity of a MS detector enabling the identification of co-migrating compounds with different m/z value.

Almost all mass analyzers can be coupled with CE (*e.g.* quadrupole, ion trap, TOF and Fourier transform ion cyclotron), whereas only few dedicated sources are commercially available. The main types are the electrospray and the atmospheric pressure chemical ionization sources: normally employed for LC/MS analysis, they could be also adapted with slight changes for CE/MS coupling. The hyphenation could be achieved by two primary approaches following described.

1. The sheath-flow interface: this approach provides the use of a supportive contact liquid into applying the voltage to the CE buffer. Two groups of liquid-supported systems are available: the sheath liquid interface and the liquid junction. The first type is characterized by a steel tube, in a coaxial placing, through the which the supportive liquid arrives to the nebulizer needle and mix with the CE buffer. This arrangement may be supported by a stream gas for the formation of the droplets. For an useful hyphenation, the capillary tip position, the flow-rate and sheath liquid composition should be optimize (Figure 4).

   In the second type the CE capillary is partially disconnected from the ESI source and filled by the buffer solution. The resultant physical and electrical disconnection allows the individual optimization of CE and ESI parameters.
2. The sheathless interface: for this approach a direct voltage application to the CE buffer is required. In order to close the electrical circuit indispensable for any CE separation, a metal coating to the end of a separation capillary should be applied. As advantage the analytes dilution by the sheath flow is avoided, on the other side the metal coating needs often to be replaced (Shamsi S.A. et al., 2004).

As previously mentioned, the CE/MS coupling has attracted considerable attention in several fields (Boone C.M. et al., 2003): toxicological analysis, environmental analysis, food analysis and also for the analysis of pharmaceutical compounds. Furthermore, although only very few dedicated commercial CE/MS are available, technological developments in instrumentation are occurring. Particularly, the introduction of bench top TOF and QTOF analyzers at affordable costs permits advance into known and unknown compounds identification through the accurate measurements of mass (Klampfl C.W., 2009).

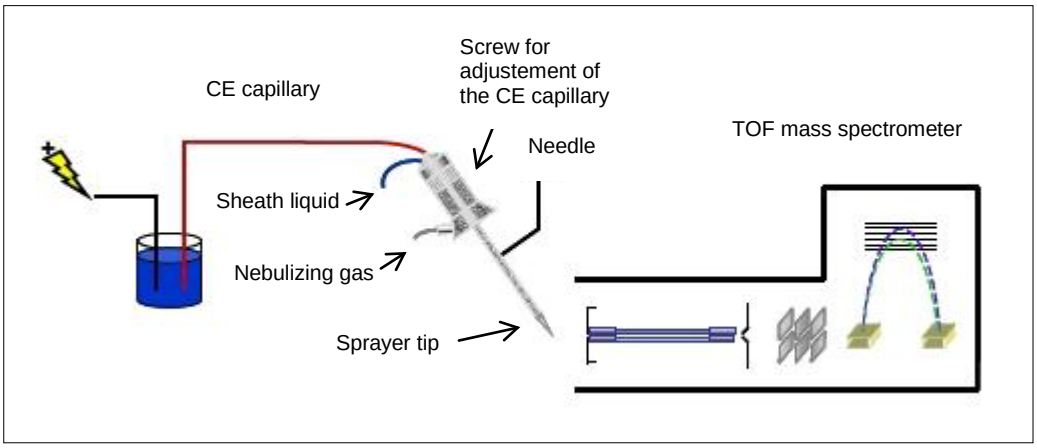


Figure 4. *Arrangement in a sheath-flow interface.*

4.2. IONIZATION

The coupling of a separation technique and a mass spectrometer is usually obtained by an interface whose role is to eliminate the solvent and produce gas-phase ions from the analyte. This can be accomplished in two ways:

- the two processes can occur in different places, *i.e.* the analyte is sampled into the vacuum region of the mass spectrometer and subsequently ionized at reduced pressure;
- ions are generated at atmospheric pressure (hence the name API) and then sampled into the vacuum region.

The second strategy has been proved to confer much more robustness and reliability to the ionization process, and as a consequence API sources have currently displaced other ion sources previously used in the routine (*e.g.* thermospray ionization [TSI]).

The main sources belonging to the family of API are listed here:

- electrospray ionization (ESI);

- atmospheric pressure chemical ionization (APCI);
- atmospheric pressure photo-ionization (APPI);
- atmospheric pressure laser ionization (APLI);
- sonic spray ionization (SSI).

Other ion sources with a limited application in routine analysis of analytical toxicology laboratories are: TSI, particle beam (PBI), fast atom bombardment (FAB) and matrix-assisted laser desorption ionization (MALDI) (Politi L. et al., 2006).

4.2.1. ELECTROSPRAY IONIZATION

In the family of API sources, ESI was the first to be designed and in the field of analytical toxicology represents **the most frequently used ionization mode**. In an ESI source a voltage (typically 3-5 kV) is applied to a capillary tube, the ESI probe. A potential difference between the ESI probe and the orifice close to the vacuum region of the mass spectrometer is applied, in order to nebulize the liquid flowing into the capillary. This happens at very low flow rates (a few microlitres per minute approximately). In an ESI source the analyte is ionized in solution. Depending on the voltage polarity, nebulized droplets trapping the ionized analyte will be positively or negatively charged. The evaporation of the solvent determine the reduction in the droplets size with a subsequent enhancement of the charge density in the droplet. When the repulsive forces between charges exceed the cohesive forces, the explosion of the droplets occurs. This reaction proceeds until the generation of gas-phase ions is obtained (Figure 5).

Typical ions obtained under ESI conditions are:

- positive mode: protonated molecular ions $[M+H]^+$, sodium or potassium adducts, solvent adducts;

- negative mode: deprotonated molecular ions $[M-H]^-$, formate or acetate adducts.

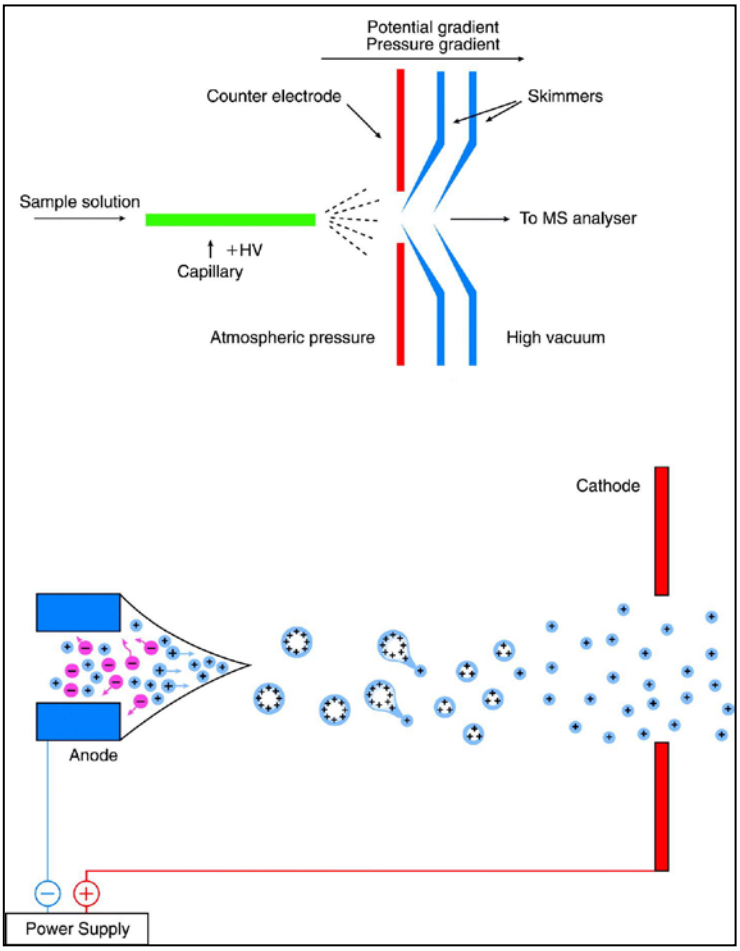


Figure 5. *Schematization of an ESI source and generation of gas-phase ions.*

It's important, especially for quantification purposes, to prevent the formation of adducts that are cause of less reproducible analysis (Law B. et al., 2000; Mortier K.A. et al., 2004; Lambert W., 2004).

ESI is a **mild process** generating a poor fragmentation. In fact, the energy transferred to the molecule during the ionization is low and this has direct influence on the production of fragment ions. As a consequence the ESI-MS does not provide much structural information. Nevertheless, the kinetic energy of the ions of the

analyte can be enhanced in order to improve the fragmentation process. This occurs within the intermediate vacuum region placed between the ion source and the mass analyzer by the application of a potential difference at the two ends of this region. Ions are accelerated and forced to collide with other molecular species (gas molecules, solvent, …) and as a consequence of these multiple collisions, the energy gained is dissipated by fragmentation. Typically, the parameter controlling this process is known as the fragmentor, orifice, skimmer or also capillary voltage (depending on the instrument manufacturer). Obviously, the higher is the voltage applied, the higher is the fragmentation rate. This technique is called **in-source collision-induced dissociation** (in-source-CID): a good strategy to obtain structural information through fragmentation, but affected by high chemical noise and the possible presence of interfering mass fragments that may complicate interpretation of a mass spectrum and identification (Politi L. et al., 2006). In fact, differently from CID obtained under MS-MS conditions, fragments detected not necessarily originate from the precursor ion.

4.2.2. ATMOSPHERIC PRESSURE CHEMICAL IONIZATION

The ionization process obtained through an APCI source is substantially based on API strategy similar to ESI, but with some peculiarities. The analyte in solution is vaporized by means of a capillary and a coaxial flow of nitrogen (nebulizer gas) into a chamber heated until 550 °C. Owing to the complete evaporation of the solvent and the analyte due to the high temperature, the ionization reaction occurs in the gas phase, contrary to ESI where ions are already present in the liquid phase. At the end of the capillary a corona discharge electrode (2-5 kV) operates as an electron source causing the ionization of the atmosphere surrounding the tip (solvent vapours,

nitrogen and oxygen). The excess of reagent ions and the analyte interact at atmospheric pressure, until ions formation through a process similar to chemical ionization (CI). The APCI source allows to carry out the ionization process both in positive and negative mode: positive ions are formed when the proton affinity of the analyte is higher than that of reagent ions, whereas negative ions are obtained when the acidity of the reagent in the gas phase is higher than that of the analyte. As the column eluate is completely evaporated in the source before ionization, APCI is suitable for the analysis of low polarity and more volatile compounds, and requires thermal stability of the analyte (Politi L. et al., 2006).

4.2.3. ATMOSPHERIC PRESSURE PHOTO-IONIZATION

Recently developed (in 2000), APPI source is quite similar to APCI, but differ from the latter for the presence of a lamp generating UV photons of a specific energy related to the gas used (xenon, argon, krypton) indispensable for the ionization of the analyte. APPI and APCI show similar application fields. However, APPI exhibits an improvement in the detection sensitivity of analytes in the low-polarity region and a higher efficiency in the ionization process by using low flow rates (< 100 μl/min) (Politi L. et al., 2006).

4.2.4. ATMOSPHERIC PRESSURE LASER-IONIZATION

APPI is a special API source based on the use of a fixed-frequency laser operating at 248 nm that allows an increasing in the photon flux during the ionization. This source constitutes an amelioration for the ionization of apolar compounds (Politi L. et al., 2006).

4.2.5. SONIC SPRAY IONIZATION

SSI allows to produce charged droplets through the flowing of a nebulizer gas (nitrogen) at sonic speed, with a subsequent formation of gas-phase ions due to the inhomogeneous charge distribution at the surface of the generated droplets. As it does not require the application of high voltage or high temperature, the main advantage of this source is related to the possibility to analyse compounds unstable under ESI conditions.

4.3. HIGH RESOLUTION MASS SPECTROMETRY

The fundamental task of a mass analyzer is to detect and separate all the ions generated from the ion source, on the basis of their different m/z value. The ability of the mass analyzer to separate ions with close m/z values is expressed as **mass resolution**. This parameter has important influence on the quality of the ions separation: the higher is the resolution of the instrument and the better is the separation of the ions in the produced mass spectrum, and an improvement in selectivity is obtained. In particular, mass resolution (R) is defined as the mass-charge value of the target peak (m/z) divided by its full width at half of its maximum height (Δm/z):

$$R_{FWHM} = m/z \,/\, \Delta m/z$$

Therefore, mass resolution varies with mass (increases at increasing masses) and is inversely proportional to peak width.

In general, mass analyzers may be classified in two groups (Holcapek M. et al., 2008):

1. low-resolution mass analyzers: resolution in the low-thousand range (*e.g.* quadrupoles and ion trap);
2. high-resolution mass analyzer: affording a minimum resolution of 15.000.

In high resolution mass spectrometry, **mass accuracy** represents another extremely important parameter to consider. Mass accuracy is usually defined as the ratio between the measured mass and the theoretical mass calculated from the elemental composition of the ion (expressed in atomic mass units [amu] or in parts per million [ppm]). High resolution mass spectrometers should allow a mass accuracy of 5 ppm or better (Holcapek M. et al., 2008). This property allows the determination of the elemental formula of small ions among several possibilities at a given nominal value or at least to significantly shorten the list of possible candidate formulas. In order to achieve accurate mass measurements, the right calibration of the mass scale should be performed and the precision of the measured mass verified by using an internal or external calibration procedure.

For the identification of the elemental composition of an unknown compound, additional information may be obtained from the comparison of the experimental isotopic pattern (IP) with the patterns predicted based on the hypothesized chemical formula (CF). This parameter is expressed as **σ value**: the lower is this value and the better is the match (typically, σ values > 0.03 exclude the match) (Politi L. et al., 2006). The σ value has very important relevance in HRMS. In fact, it has been demonstrated that high mass accuracy and high resolving power are not sufficient to reduce the number of possible CF to a minimum (Kind T., 2006), and the σ value (and even the isotopic pattern spacing) become a fundamental requirement for further to achieve the required selectivity.

Recently, TOF analyzers have found wide diffusion for the screening of unknown compounds providing high mass accuracy at affordable costs (Nielen M.W.F. et al., 2007). However, mass analyzers based on Ion Cyclotron resonance and Fourier transformation are utilized for analogous purposes providing even better mass accuracies (Schrader W. et al., 2004; Gratz S.R. et al., 2006; Vu H. et al., 2008), but their diffusion is limited cause of the high costs and special technical demands for installation and operation (*e.g.* the ion cyclotron resonance [ICR] and the newest Orbitrap mass analyzer). In this context, the selectivity power of high resolution mass spectrometry (providing the accurate determination of mass and sotopic pattern) combined with high efficiency separation technique, shows interesting perspectives for the development of new and promising strategies of analysis, particularly in the field of general unknown screening.

4.3.1. ION CYCLOTRON RESONANCE MASS ANALYZER

The FTICR is essentially a mass analyzer and a detector at the same time. The basic rule of an ICR mass analyzer consists in the motion of a charged particle in a cell submitted to a very high magnetic field. This field force ions to move through a circular trajectory in a plain perpendicular to the magnetic field (cyclotron motion). The ions move between opposite electrode plates causing the production of a small alternating electric current. The amplitude of the current depends on the ions number, which makes quantitation possible, the frequency is that of the ions rotation in the magnetic field (cyclotron frequency). Consequently, the detection occurs without collision of the ions, allowing the repeated measurement of a single ion and resulting in an increase of the sensitivity. The trapped ions become excited by applying radio-frequency electric potentials. The cyclotron frequency is inversely

correlated to the m/z value of the ions, as a consequence the recorded frequency is the sum of the frequencies of the several ions at different m/z value. Therefore, the mass spectrum is calculated by the mathematical extraction (Fourier Transformation) of the frequencies and the abundances corresponding to different ions. In order to avoid ions collision, the FTICR is maintained at very high vacuum (10^{-8} – 10^{-9} torr) (Hendrickson C.L. et al., 1999).

4.3.2. TIME-OF-FLIGHT MASS ANALYZER

TOF instrumentation is based on the fundamental principle that different ions moving in the same direction at constant kinetic energy, will have a velocity inversely proportional to their m/z ratio. Usually, the flight time for this type of analyzer is in the range of 5-100 μs. An important task for obtaining the accurate measurement of the flight time is that ions generated from the source, enter the flight tube in a very brief and precise interval, so that their kinetic energy is not too spread.

A TOF analyzer typically used in a routine laboratory has a 1-m long fight distance.

In TOF mass analysis, several phenomena should be considered because of their impact on mass resolution.

1. Temporal distribution: ions with identical m/z value may be generated with the same kinetic energy but at different times, thus crossing the mass analyzer with a steady time difference and a steady space difference.
2. Spatial distribution: ions with identical m/z value may be generated at the same time with identical initial kinetic energy, but at different locations within the source. Consequently, the ions at the back of the source will be submitted to higher potential gradient producing an acceleration at higher kinetic energy with respect to that of the ions formed in other places.

3. Kinetic energy distribution: ions with different initial kinetic energy could be generated. Particularly, ions will be normally accelerated, but the final velocities will be different and isobaric ions will arrive at unequal times to the detector. A similar situation occurs also for ions with identical kinetic energy but velocities in various trajectories. Furthermore, ions initially flapping to the detector arrive before ions initially flapping away from it. In order to arrest the latter ions, a deceleration is induced before the subsequent reacceleration. This approach enables the ions to return to their original position. The previously mentionated phenomenon is well known as "turn-around time" (Staub A. et al., 2009).

Recently, technological advances have provided the possibility to partially overcome many of these limitations, *i.e.* the **reflectron** and the **orthogonal acceleration**. The reflectron is primary constituted of one or more retarding fields (decelerating and reflecting fields) with an opposite orientation with respect to the accelerating field. The ions with identical m/z value, but different kinetic energies, enter in the field at different degrees of path. Obviously, ions with higher velocity penetrate the decelerating field more deeply than ions with lower velocity and vice versa. The main characteristic of a TOF analyzer with the orthogonal acceleration is that the ionization source, the accelerator and the detector form an angle slightly superior to 90° (Figure 6). This feature provides several advantages like minimal temporal dispersion and simultaneous correction of the spatial and velocity distributions.

A TOF analyzer provided with reflectron and orthogonal acceleration technology, allows to perform analysis at very high mass resolution (resolving power up to 10.000) and mass accuracies better than 2-5 ppm at affordable costs.

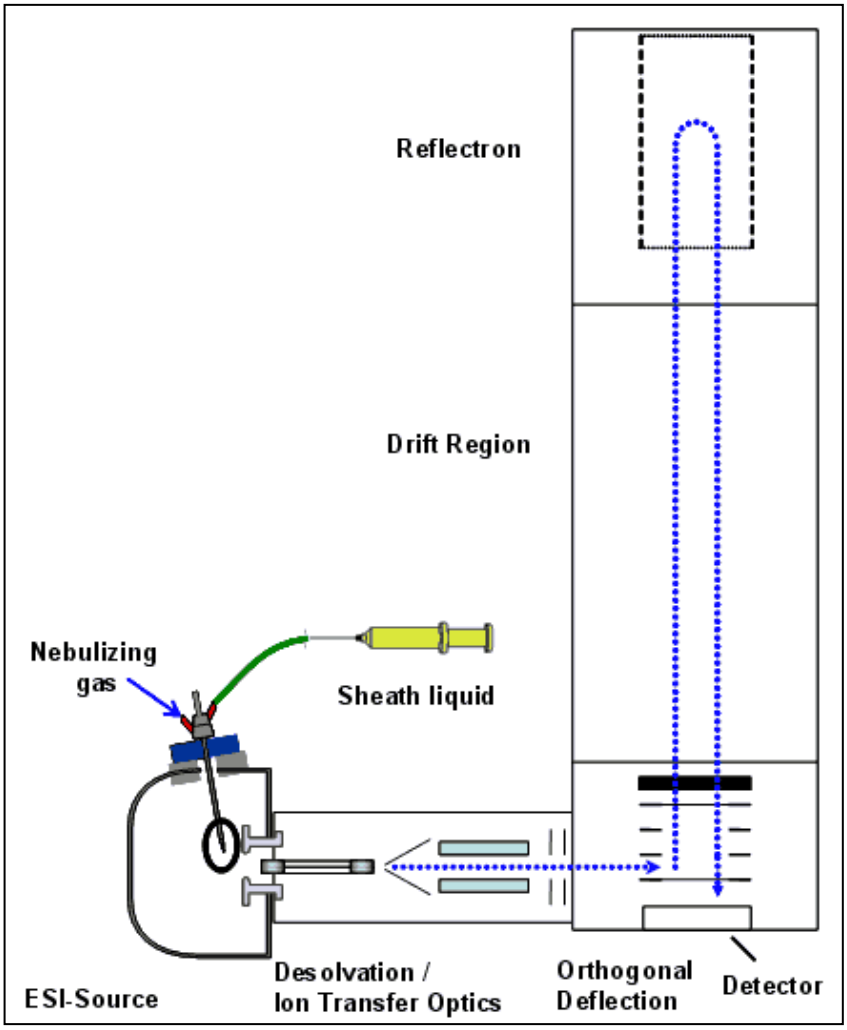


Figure 6. *Orthogonal ESI-TOF scheme*

5. EXPERIMENTAL

The general unknown screening has strongly benefited from the coupling of API sources with high resolution mass spectrometry (HRMS) detection. As illustrated earlier, the accurate measurement of mass and IP allow the determination of the elemental composition of a given unknown compound. In contrast to the so-called library search approach, by using API/HRMS the construction of expensive, time consuming and thus limited mass spectral reference databases, may be avoided. However, there is no biunique relation between chemical formula and structural formula and a database of compound names/associated molecular formulas is required.

To this purpose, recently, *we have created a **database** containing more than 50.000 PTRC (PTRC Database) and their phase I and II metabolite*s. Basically, it consists in a subset of the PubChem Compounds database (www.pubchem.ncbi.nlm.nih.gov, 2007). We have demonstrated that the use of this tool enables a wider screening with respect to databases including few hundreds of reference substances.

However, several compounds may have the same elemental composition and the number of isobars increases with the dimension of the reference compounds database. As a consequence, the search in a large database as the PTRC may result in a list of compounds with identical chemical formula. In order to discriminate between isobaric compounds, additional tools are necessary.

In pursuit of this goal, *we have developed two new data processing strategies that may also be combined: a **"metabolomic" approach**, based on the detection of a*

*predetermined list of metabolites and a **chemometric approach**, based on the estimation of the RPLC retention behaviour of each candidate from its molecular descriptors as calculated from its structure.*

6. CHEMICAL FORMULAS DATABASE FOR THE SCREENING OF PTRC

The detection/identification of unknown pharmaco/toxicologically relevant compounds (PTRC) in biosamples represents a hard challenge for analytical toxicologists. The extremely wide range of PTRC - in terms of molecular weight (MW), polarity, pKa, chemical/thermal stability - on one side, and the high number/amount of interferences in the substrate on the other make this goal similar to searching for the needle in the haystack.

Yet, general unknown search of PTRC has important clinical/forensic applications: from diagnosis of acute/lethal intoxication, to identification of drugs affecting performance at work (workplace drug testing), while driving (driving under influence of drugs), and in sports (doping control), from applications in transplantation surgery (excluding drugs presence in the organ's donor) to screening of xenobiotics in foodstuffs/environmental matrices (Hogendoorn E.A. et al., 1996; De Zeeuw R.A., 1998; Mottram D.R., 1999; Polettini A., 1999).

In the past, PTRC screening was mainly based on gas chromatography-mass spectrometry (GC-MS) (Polettini A., 1998; Maurer H.H., 2004). In fact, GC separation power and MS detection selectivity were found to largely counterbalance the pitfalls of GC, *i.e.* the need to isolate analytes from the aqueous substrate and to derivatize polar compounds (Polettini A., 1998). Moreover, the inherent simplicity of information of a mass spectrum and the availability of a well established and standardized ionization technique (electron ionization), allowed the build-up of large databases of reference mass spectra (http://www.nist.gov, 2007; http//eu.wiley.com, 2007), some

specifically devoted to PTRC (Maurer H.H., 2007), and development of fast computer aided identification based on library search.

At the beginning of 1990s the advent of atmospheric pressure interfaces (API) brought about ruggedness in coupling liquid chromatography (LC) to MS, giving new impulse to research in PTRC screening. The much higher versatility of LC (suitable for polar and thermally labile compounds) indeed appeared to enable widening the range of identifiable compounds (Polettini A., 2006). Different research groups investigated the library-search based LC-MS approach through the set-up of PTRC spectral libraries (Weinmann W., 1999; Marquet P., 2002; Gergov M., 2003). However, obstacles were found to hamper development of broad-range procedures as previously had happened with GC-MS. First, API techniques provide soft ionization, accounting for little fragmentation of pseudomolecular ion and poor structural information (Politi L., 2006). Increased fragmentation is obtainable through in-source collisionally induced dissociation (CID), or by means of tandem mass spectrometric (MS-MS) techniques. However, both approaches have limitations. In the first, spectra may be contaminated by chemical noise whereas in the latter parent ions submitted to CID have to be selected in advance, thus limiting the range of compounds identifiable (Gergov M., 2003). The recent implementation of data-dependent acquisition routines where all precursor ions above a certain abundance threshold are submitted to CID, overcomes this limitation of MS-MS screening applications (Sauvage F.L. et al., 2006). Another obstacle, however, is the scarce reproducibility of API full scan and product ion spectra at interlaboratory level (Jansen R. et al., 2005). Although tuning procedures have been proposed (Weinmann W. et al., 2001), this problem has limited thus far the development of spectral databases as large as those used in GC-MS. Currently, LC-MS libraries

contain spectra of up to twelve hundreds compounds (Sauvage F.L. et al., 2006; Dresen S. et al., 2006).

Recently, bench-top time-of-flight (TOF) mass spectrometers have become available as GC, LC, and capillary electrophoresis (CE) detectors opening up new perspectives in PTRC screening. TOF analysers enable acquisition at high mass resolution and provide mass accuracies of few ppm. This feature, combined with the accurate determination of the isotopic pattern (IP), allows unequivocal identification of the elemental composition of small ions (m/z 200-500) (Pelzing M. et al., 2004). Studies carried out so far with LC-TOF have demonstrated the suitability of this screening approach in the fields of clinical/forensic toxicology and pesticide residue analysis (Fang L. et al., 2003; Ibanez M. et al., 2005; Ojanpera S. et al., 2006; Sancho J.V. et al., 2006; Kolmonen M. et al., 2007), its considerable advantage being to avoid building-up mass spectral libraries. Screening procedures based on chemical formula (CF) identification require a simple reference database consisting of a listing of compound names/associated molecular formulas. In fact, high resolution MS software enables search for CFs whose calculated mass/IP matches with that measured for the unknown, but do not provide hints on structure and compound's name. Either in-house (Ojanpera S. et al., 2006; Kolmonen M. et al., 2007) or commercially available (Ibanez M. et al., 2005) databases have been used so far for this purpose.

Ojanpera and colleagues created an in-house database of 735 PTRC including therapeutic and illicit drugs, designer drugs, and metabolites (Ojanpera S. et al., 2006. Subsequently, they developed another database of 124 doping agents (70 metabolites) (Kolmonen M. et al., 2007).

Commercial databases containing CFs of PTRC are available. For example, the Merck Index (O'Neil M. et al., 2006) lists over 10000 different compounds, but no metabolites. Metabolites identification is essential in analytical toxicology, especially for compounds undergoing extensive biotransformation. Another commercial database is the MPW library (Maurer H. et al., 2007) listing over 7800 PTRC, including metabolites, but this is GC-oriented and, therefore, of limited utility in a TOF-MS screening approach, unless based on GC. Yet, it is recognized that the higher versatility of LC and CE make these separation techniques better suited, in combination with MS, for screening purposes (Polettini A., 2006; Gottardo R. et al., 2007).

The NIST Chemistry Webbook (http://webbook.nist.gov/chemistry/, 2007) database contains about 70000 compounds. It is freely accessible and searchable through the web, contains some metabolites (although not glucuronides), but includes also non-PTRC compounds (*e.g.* petroleum derivatives, cosmetics, etc.).

PubChem Compound (www.pubchem.ncbi.nih.gov, 2007), a component of U.S. National Institutes of Health's Molecular Libraries Roadmap Initiative, is a large database of biologically-active small molecules. It contains over 10 millions records collected from different sources (*e.g.* EPA, Cambridge Soft, Discovery Gate, ChemIDPlus, etc.) - which makes it obviously too large for an effective general unknown screening. However, subsets can be created: for example, 270.000 compounds are listed under the category "toxicology". PubChem is freely searchable through the web and also freely downloadable as a text file and converted into a standalone searchable database.

In the frame of the setup of a "general unknown search" procedure based on CF identification, our purpose was the implementation of a reference database of

compound names/CFs/exact masses, based on a subset of PubChem Compound database specifically devoted to PTRC, metabolites and related substances and to test its performance in identifying unknown PTRC in biological samples.

6.1. MATERIALS AND METHODS

6.1.1. CHEMICAL AND STANDARDS

All solvent and reagents (Carlo Erba, Milan, Italy) were HPLC or analytical grade. The running electrolyte solution was 25 mM ammonium formate, adjusted to pH 9.5 with ammonia. Solvent extraction was carried out using Toxi-Tubes A (Varian, Lake Forest, CA, USA). Analytical standards were from Salars (Como, Italy).

6.1.2. SAMPLE COLLECTION AND PREPARATION

Hair was decontaminated with an aqueous solution of 0.3% Tween-20, cut into small fragments and incubated overnight in 0.1 M HCl (1 ml/45 °C). Incubation mixtures were neutralized with NaOH and extracted with Toxi-Tubes A. Protein precipitation was used for cadaveric **blood**. A 1-ml aliquot of saturated $(NH_4)_2SO_4$ solution was slowly added to 1-ml blood; after 10 minutes the mixture was centrifuged (3500 rpm, 10 min); the supernatant (1 ml) was mixed with 1 ml of 0.1 M Na_2HPO4 (pH 8.9) and extracted with chloroform-2-propanol (9:1, 5 ml) by vortex-mixing (2 min). The organic phase (4 ml) was evaporated to dryness and the residue dissolved in water (250 µl) (Tagliaro F. et al., 1994). **Urine** was submitted to dilution (1:5 with water) and centrifugation.

6.1.3. CE/MS INSTRUMENTATION

TOF analyses were carried out using a MicrOTOF mass spectrometer (Bruker Daltonics, Bremen, Germany) coupled to a P/ACE 5500 automated capillary electropherograph (Beckman, Fullerton, CA, USA) with an electrospray (ESI) source. Untreated fused-silica capillaries (75 µm ID, 100 cm length, Composite Metal Services, The Chase, Hallow, Worcestershire, UK) were used, directly connected to the ESI needle at their cathodic end. The following parameters of CE–MS coupling were optimized as previously reported (Manetto G. et al., 2000): separation voltage, buffer composition/pH, injection mode, sheath liquid flowrate and composition.

Separation voltage was 15 kV (forward polarity). The cooling liquid was kept at 20 °C. Field-amplified sample stacking (FASS) injections were carried out. The cathode end of capillary was connected to the orthogonal ESI source using a coaxial sheath-liquid interface (Agilent Technologies, Palo Alto, CA, USA) positioned orthogonally to the source, the spray needle being grounded. Capillary voltage was -4 kV, source temperature 200 °C. Nitrogen was used as drying and nebulizing gas (drying gas flow rate: 5 l/min, nebulizer pressure: 60 kPa). The MS was operated in positive ion (50-750 *m/z*; acquisition scan rate, 20 spectra/s). External calibration was obtained by infusing for 1 min at the beginning of each run a mixture of 10 mM NaOH in 2-propanol and 0.2% formic acid (1:1, v/v), using seven calibration ions corresponding to the formulas Na(NaCOOH)x, with x= 2 to 9. The nominal resolution of the instrument was 10000 (Full Width at Half Minimum), accuracy 5 ppm. A coaxial sheath liquid consisting of a mixture of 2-propanol/water (50/50) added with 0.5% formic acid was delivered at 4 µl/min by a syringe pump (Cole-Parmer, Vernon Hill, IL, USA). Data processing was carried using Data Analysis software (Version 3.2, Bruker Daltonics) and Microsoft Excel (2002 version) as follows. The base mass

peak (after background subtraction) measured for a given unknown electrophoretic peak was searched, after proton subtraction, into Pubchem Compound PTRC subset database (±10 ppm tolerance). For each retrieved CF, mass error (difference between measured and theoretical mass), and sigma (a parameter, calculated by the Bruker software, accounting for the difference between theoretical and measured IP; the lower the value, the better the matching) were calculated.

6.1.4. THE PUBCHEM COMPOUND PTRC SUBSET DATABASE

In order to remove from PubChem Compound as much non-relevant compounds as possible and to reduce its size, a subset was created by applying the following limitations:

- compounds listed in the category “toxicology”;
- source: ChemIDPlus (http//chemsis.nlm.nih.gov/chemidplus/, 2007) ;
- compounds with linking information to records in the National Library of Medicine's Medical Subject Heading (MeSH) database (http//www.nlm.nih.gov/mesh/mBrowser.html, 2007);
- MW: 100-750;
- elements in CF: C,H,O,N,F,Cl,Br,I,P,S.

Each limit was adopted upon verification that size reduction did not compromise comprehensiveness. The subset so defined contains about 50,500 compounds. Data listed for each compound include common and, often, IUPAC names, CF and other calculated chemical/physical parameters. The subset (as defined by setting the criteria listed above in the “Limits” webpage of PubChem) was retrieved both in “Summary” and “Property report” format. Each format was then downloaded as text file and, after removing variable names and special characters, converted into Excel

format. The two data sets (“Summary” and “Property report”) were then joined. Finally, by using an home-made algorithm able to extract from the CF the number of atoms of each element, the exact mass of each CF was calculated.

6.2. RESULTS AND DISCUSSION

6.2.1. FEATURES OF PUBCHEM COMPOUND PTRC SUBSET DATABASE

One interesting feature of Pubchem Compound PTRC subset database is comprehensiveness. The database is rich not only in parent compounds but also in metabolites (about 6000 phase I metabolites and 180 glucuronides), precursors, and other compounds related to a parent drug (*e.g.* drug esters), as well as in toxicologically relevant endogenous compounds (*e.g.* GHB, testosterone). In order to test the database for comprehensiveness, a comparison was made with the MPW (Maurer H.H., 2007) mass spectral database within the MW range 250-350: over 95% of the parent compounds with mass spectral information listed in the latter are represented in the Pubchem subset, which by converse contains about 5 times as many compounds as the MPW mass spectral database.

Table 1 shows some examples: 17 cocaine-related compounds are present, and 30 records for heroin, including glucuronides of morphine and codeine, and even glutathionyl-morphine. Other examples of glucuronides are shown in Table 1.

In contrast, few nor- metabolites have not been found (norhydrocodone; noralprenolol; norlevomepromazine sulfoxide), though the relative parent drug and one or more other metabolites are usually present.

Table 1. *Examples of groups of related compounds included in the database.*

Compound Name 1	Chemical Formula	Theoretical Mass
Anhydroecgonine, Ecgonidine	C9H13NO2	167.094629
Methylecgonidine, Anhydromethylecgonine	C10H15NO2	181.110279
Ecgonine, Ekgonin	C9H15NO3	185.105193
Methyl ecgonine, Methylecgonine	C10H17NO3	199.120843
Ethylecgonine, Ethyl ecgonine	C11H19NO3	213.136493
Nortropacocaine, O-Benzoylnortropine	C14H17NO2	231.125929
Benzoylnorecgonine, O-Benzoylnorecgonine	C15H17NO4	275.115758
N-Benzoylnormethylecgonine	C16H19NO4	289.131408
Norcocaine	C16H19NO4	289.131408
benzoylecgonine, O-Benzoylecgonine	C16H19NO4	289.131408
Norcocaethylene, Benzoylnorecgonine ethyl ester	C17H21NO4	303.147058
Pseudococaine, Depsococaine	C17H21NO4	303.147058
cocaine, Eritroxilina	C17H21NO4	303.147058
N-Hydroxynorcocaine, 8-Hydroxynorcocaine	C16H19NO5	305.126323
m-Hydroxybenzoylecgonine, 3-Hydroxybenzoylecgonine	C16H19NO5	305.126323
3'-Hydroxycocaine, 3'-Hydroxybenzoylecgonine methyl ester	C17H21NO5	319.141973
Cinnamoylcocaine, 521-67-5	C19H23NO4	329.162708
Normorphine, Demethylmorphine	C16H17NO3	271.120843
Pseudomorphine, gamma-Isomorphine	C17H19NO3	285.136493
Norcodeine, N-Norcodeine	C17H19NO3	285.136493
morphine, Morphium	C17H19NO3	285.136493
codeine, Methylmorphine	C18H21NO3	299.152144
Morphine-7,8-oxide, Morphinan-3,6-diol	C17H19NO4	301.131408
10-Hydroxymorphine, 10 alpha-Hydroxymorphine	C17H19NO4	301.131408
Paramorphine, Thebain	C19H21NO3	311.152144
Codeine epoxide, Codeine-7,8-oxide	C18H21NO4	315.147058
14-Hydroxycodeine, Morphinan-6,14-diol	C18H21NO4	315.147058
codeine, Brontex	C18H23NO4	317.162708
Palaudine, 3'-Hydroxypapaverine	C19H19NO4	325.131408
6-Hydroxypapaverine	C19H19NO4	325.131408
4'-Hydroxypapaverine	C19H19NO4	325.131408
3-Acetylmorphine, O3-Acetylmorphine	C19H21NO4	327.147058
6-Acetylmorphine, Monoacetylmorphine	C19H21NO4	327.147058
3-Acetylmorphine, Monoacetylmorphine	C19H21NO4	327.147058
papaverine, Papaverine	C20H21NO4	339.147058
Acetylcodeine, Codeine, acetyl-	C20H23NO4	341.162708
Diacetylmorphine, Acetomorfine	C21H23NO5	369.157623
Pavabid, papaverine	C20H22ClNO4	375.123736
noscapine, Narcompren	C22H23NO7	413.147452
noscapine, Gnoscopine	C22H23NO7	413.147452
Normorphine-6-glucuronide	C22H25NO9	447.152931
Normorphine-3-glucuronide	C22H25NO9	447.152931
noscapine, Narcotussin	C22H24ClNO7	449.124130

Morphine-3-glucuronide	C23H27NO9	461.168581
Morphine-6-glucuronide	C23H27NO9	461.168581
Codeine-6-glucuronide, 20736-11-2	C24H29NO9	475.184232
Morphine-gsh, 10-alpha-S-Glutathionylmorphine	C27H34N4O9S	590.204649
3-Hydroxyflunitrazepam, DP 327	C16H12FN3O4	329.081184
N-Demethylflunitrazepam, N-Desmethylflunitrazepam	C15H11FN2O	254.085541
7-Aminonorflunitrazepam, 7-Amino-desmethylflunitrazepam	C15H12FN3O	269.096440
7-Aminoflunitrazepam, 7-Amino-flunitrazepam	C16H14FN3O	283.112090
flunitrazepam, Rohypnol	C16H12FN3O3	313.086269
lorazepam, Ativan	C15H10Cl2N2O2	320.011933
N-(2-Hydroxyethyl)lorazepam, SAS 632	C17H14Cl2N2O3	364.038148
Lorazepam glucuronide, 32781-79-6	C21H18Cl2N2O8	496.044021
furosemide, Frusemide	C12H11ClN2O5S	330.007720
Furosemide glucuronate, Furosemide glucuronide	C18H19ClN2O11S	506.039808
(Dideacetoxy)pancuronium, (Didesacetoxy)pancuronium	C31H56Br2N2	614.281025
3-(Deacetoxy)pancuronium, 3-(Desacetoxy)pancuronium	C33H58Br2N2O2	672.286504
Pavulon, pancuronium	C35H60Br2N2O4	730.291984
Norbuprenorphine, N-Desalkylbuprenorphine	C25H35NO4	413.256609
Demethoxybuprenorphine	C28H37NO3	435.277344
buprenorphine, Buprenex	C29H41NO4	467.303559

Other metabolites not included are EMDP (though methadone, EDDP, methadol, p-hydroxymethadone, and normethadone are present), and NAP 226-90 (rivastigmine metabolite) (Pommier F. et al., 2003). The database contains some replicates (*e.g.* 3-acetylmorphine and noscapine in Table 1) and a number of errors (*e.g.* the second hit for papaverine in Table 1, which is the hydrochloride instead).

The other side of the coin of comprehensiveness is that the probability that more than one hit falls within the defined mass tolerance or, worse, shares the same CF is much higher than in a 1000-compounds database. Discrimination between different CFs can be usually accomplished by comparing the isotopic peaks pattern (from M+1 up to M+4) with the unknown. Ojanpera et al. found that IP matching provides a powerful means of identification in combination with accurate mass (Ojanpera S. et

al., 2006). However, this approach does not help when different compounds share the same CF. The average number of hits with identical CF in the database is 1.82±2.27 (median=1; range 1-39), typically decreasing at the increase of mass.

6.2.2. PERFORMANCE EVALUATION OF PUBCHEM COMPOUND PTRC SUBSET DATABASE

The performance of Pubchem Compound PTRC subset database in identifying unknowns was evaluated by analysing biosamples collected from subjects under treatment with known drugs or by comparison with reference standards (86 cases examined so far). Selected examples are illustrated in the following.

Example 1 concerns identification of benzoylecgonine in blood of a cocaine-overdose case (Table 2 - example 1). The sample previously tested positive for cocaine metabolite by routine immunoassay. Only one CF matched with measured mass of the peak (*m/z* 290.1387): $C_{16}H_{19}NO_4$. However, this pertains to 7 compounds in the database, 3 of which are cocaine related compounds (N-benzoylnormethylecgonine, norcocaine and benzoylecgonine). Based on the identification of cocaine and benzoylecgonine by comparison of migration times, mass, and spectral data with pure standards (fragment at *m/z* 182.1179 and *m/z* 168.1038 in the spectrum of cocaine and benzoylecgonine, respectively, both corresponding to loss of $C_7H_5O_2$) other cocaine-related compounds were tentatively identified in the sample: ecgonine, methylecgonine, ethylecgonine, norcocaine, hydroxynorcocaine , cocaethylene, hydroxycocaine (p and m, partially overlapped), cynnamoylcocaine (cis and trans) (Figure 7).

Example 2 illustrates identification of citalopram and its desmethyl metabolite in hair of a subject under therapy with escitalopram (Table 2 – example 2). Parent

compound exhibited a peak at m/z 325.1714. The search in the database retrieved CF of citalopram/escitalopram ($C_{20}H_{21}FN_2O$) and another formula ($C_{23}H_{20}N_2$). However, the IP matching of the latter with the unknown was worse (sigma, 0.0458 and 0.0271, respectively). Identification of citalopram/escitalopram was confirmed by the presence of desmethylcitalopram in the sample (peak at m/z 311.1551). Four different CFs were retrieved in this case, again that of desmethylcitalopram had the best matching (lower sigma). Didesmethylcitalopram, though present in the database, was not detected in the sample.

Examples 3 and 4 show identification of chlorpromazine and metabolites in urine of a woman under therapy who committed suicide by hanging (Table 2 - example 3, Figure 8) and of tramadol and metabolites in urine of a treated patient (Table 2 – example 4, Figure 9). In both cases, together with parent compound, a number of metabolites where tentatively identified. Although for all analytes more than one hit is retrieved from the database, the simultaneous presence of electrophoretic peaks at m/z compatible with that of different metabolites of a unique parent drug strongly support identification.

Table 2. *List of candidates for the examples of identifications illustrated in the text.*

Example	Compound Name 1	Chemical Formula	Theoretical Mass	Mass Error (ppm)	Sigma
1	Dihydrolycorine, Lycorine, dihydro-	C16H19NO4	289.131408	-0.179	0.0234
	Pseudolycorine, psi-Lycorine	C16H19NO4	289.131408		
	N-Benzoylnormethylecgonine	C16H19NO4	289.131408		
	Zephyranthine, 2030-55-9	C16H19NO4	289.131408		
	Dadmcep, 2,6-Diacetyl-1,5-dimethyl-7-(2-carboxyethyl)-3H-pyrrolizine	C16H19NO4	289.131408		
	Norcocaine	C16H19NO4	289.131408		
	benzoylecgonine, O-benzoylecgonine	C16H19NO4	289.131408		

2	N-NCP, N-Cyclohexyl-N'-(1-pyrenyl)carbodiimide ...	C23H20N2	324.1626487	-4.398	0.0458
	Escitalopram, (S)-Citalopram	C20H21FN2O	324.1637915	-0.884	0.0271
	citalopram, Nitalapram	C20H21FN2O	324.1637915		
	UK 1745, UK-1745 ...	C16H23ClN2O2	310.1448057	-9.641	0.1593
	Isamoltane, Cgp 361° ...	C16H23ClN2O2	310.1448057		
	Alloclamida, Alloclamide ...	C16H23ClN2O2	310.1448057		
	bifonazole, Mycospor ...	C22H18N2	310.1469986	-2.593	0.0454
	Monodesmethylcitalopram	C19H19FN2O	310.1481414	1.080	0.0270
	9-Daeta, 9-Diethylaminoethylthioacridine	C19H22N2S	310.1503694	8.240	0.0349
	Nothiazine, Meprazine	C19H22N2S	310.1503694		
3	Thiopyronine, 3,6-Bis(dimethylamino) thioxanthylium chloride	C17H19ClN2S	318.095747	-4.959	0.0104
	Tannoflavine T, Setoflavine T	C17H19ClN2S	318.095747		
	chlorpromazine, Propaphenin	C17H19ClN2S	318.095747		
	6-(4-Vinylbenzyl-n-propyl)amino-1,3,5-triazine-2,4-dithiol	C15H18N4S2	318.097288	4.258	0.1322
	2-dimethylamino-1-phenothiazin-10-yl-propan-1-one hydrochloride	C17H19ClN2OS	334.090662	-5.354	0.0365
	Chlorpromazine N-oxide,	C17H19ClN2OS	334.090662		
	Opromazine, chlorpromazine sulphoxide	C17H19ClN2OS	334.090662		
	8-Hydroxychlorpromazine	C17H19ClN2OS	334.090662		
	7-Hydroxychlorpromazine	C17H19ClN2OS	334.090662		
	Moptfeq, 6-Methoxy-4-phenyl-1-(2,2,2-trifluoroethyl)-2(1H)-quinazolinone	C17H13F3N2O2	334.092912	1.362	0.1805
	Triflubazam, Triflubazam	C17H13F3N2O2	334.092912		
	N-Desmethylchlorpromazine	C16H17ClN2S	304.080097	-9.185	0.0302
	Chlorfenethazine, Chlorphenethazine	C16H17ClN2S	304.080097		
	p-Nitrophenyl-tma-carbonate, p-Nitrophenyl-tma-carbonate chloride	C12H17ClN2O5	304.082599	-0.983	0.0475
	Xanthocillin Y 1, BRN 4159253	C18H12N2O3	304.084792	6.205	0.2269
	Abbott 79175	C15H13FN2O4	304.085935	9.951	0.229
	RS-86, 2-Ethyl-8-methyl-2,8-diazaspiro(4,5)decane-1,3-dione hydrobromide	C11H19BrN2O2	290.062991	-6.896	0.268
	Didemethylchlorpromazine	C15H15ClN2S	290.064447	-1.892	0.0193
	Fudiolan, Fusione	C12H18O4S2	290.064650	-1.193	0.1471
	CEDU, CCRIS 2828	C11H15ClN2O5	290.066949	6.705	0.0476
	BRN 0820355, 5-24-05-00358 (Beilstein Handbook Reference)	C12H10F4N2O2	290.067840	9.767	0.2213
	Didesmethylchlorpromazine sulfoxide	C15H15ClN2OS	306.059362	4.195	0.0313
	NSC90450, EINECS 222	C12H10N4O6	306.060034	2.005	0.2122
4	5-Meo-dpac, 5-Methoxy-3-(dipropylamino)chroman	C16H25NO2	263.1885291	1.857	0.0043
	Dendrobine	C16H25NO2	263.1885291		
	6,7-Dihydroxy-2-N,N-dipropylaminotetralin	C16H25NO2	263.1885291		

Desvenlafaxine, Norvenlafaxine	C16H25NO2	263.1885291		
N,N-Dipropyl-5,6-adtn	C16H25NO2	263.1885291		
2-Ppmm, 2-n-Pentyloxy-2-phenyl-4-methylmorpholine	C16H25NO2	263.1885291		
Tramadol, Ultram	C16H25NO2	263.1885291		
Butacarb	C16H25NO2	263.1885291		
tramadol hydrochloride	C16H25NO2	263.1885291		
Metazine	C11H19N7	249.1701936	-8.756	0.0145
Spasmocalm, Dolispan	C15H23NO2	249.172879	1.979	0.0021
N-Desmethyltramadol	C15H23NO2	249.172879		
4-Allylphenoxypropanolamine	C15H23NO2	249.172879		
9-Phenylnonanohydroxamic acid	C15H23NO2	249.172879		
Ndo 008	C15H23NO2	249.172879		
O-Desmethyltramadol	C15H23NO2	249.172879		
3-tert-Butylphenyl sec-butylcarbamate	C15H23NO2	249.172879		
3-((Dimethylamino)(2-hydroxycyclohexyl)methyl)phenol	C15H23NO2	249.172879		
2-Propanol, 1-(4-indanyloxy)-3-(isopropylamino)-	C15H23NO2	249.172879		
6-Hddpb, DP-6OH-3CA	C15H23NO2	249.172879		
8-Hddpb, DP-8OH-3CA	C15H23NO2	249.172879		
Procinolol	C15H23NO2	249.172879		
Ciramadol	C15H23NO2	249.172879		
alprenolol	C15H23NO2	249.172879		

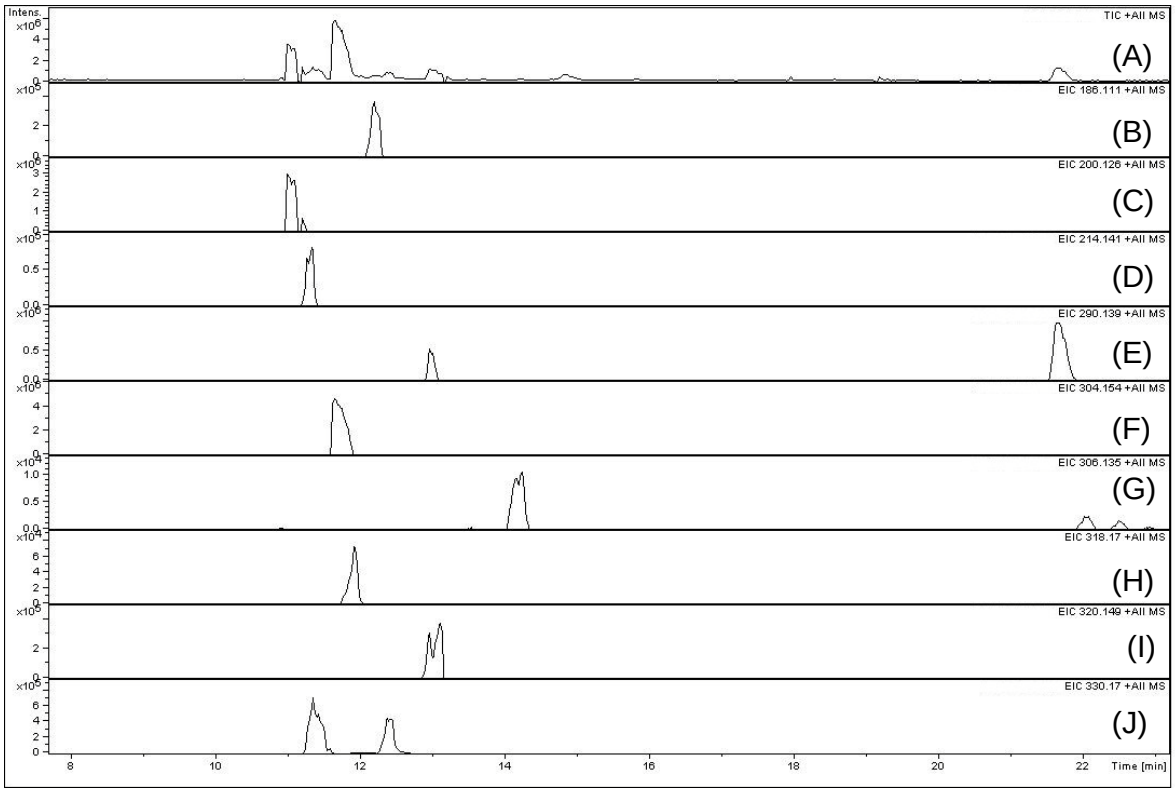


Figure 7. *Total Ion electropherogram (A) and mass electropherograms of cocaine metabolites and related compounds tentatively identified in cadaveric blood of a cocaine-overdose case through the search in the Pubchem Compound PTRC subset database: m/z 186.111 (B, ecgonine), m/z 200.126 (C, methylecgonine), m/z 214.141 (D, ethylecgonine), m/z 290.139 (E, norcocaine and benzoylecgonine), m/z 304.154 (F, cocaine), m/z 306.135 (G, hydroxynorcocaine), m/z 318.170 (H, cocaethylene), m/z 320.149 (I, p- and m-hydroxycocaine), m/z 330.170 (J, cis and trans cynnamoylcocaine).*

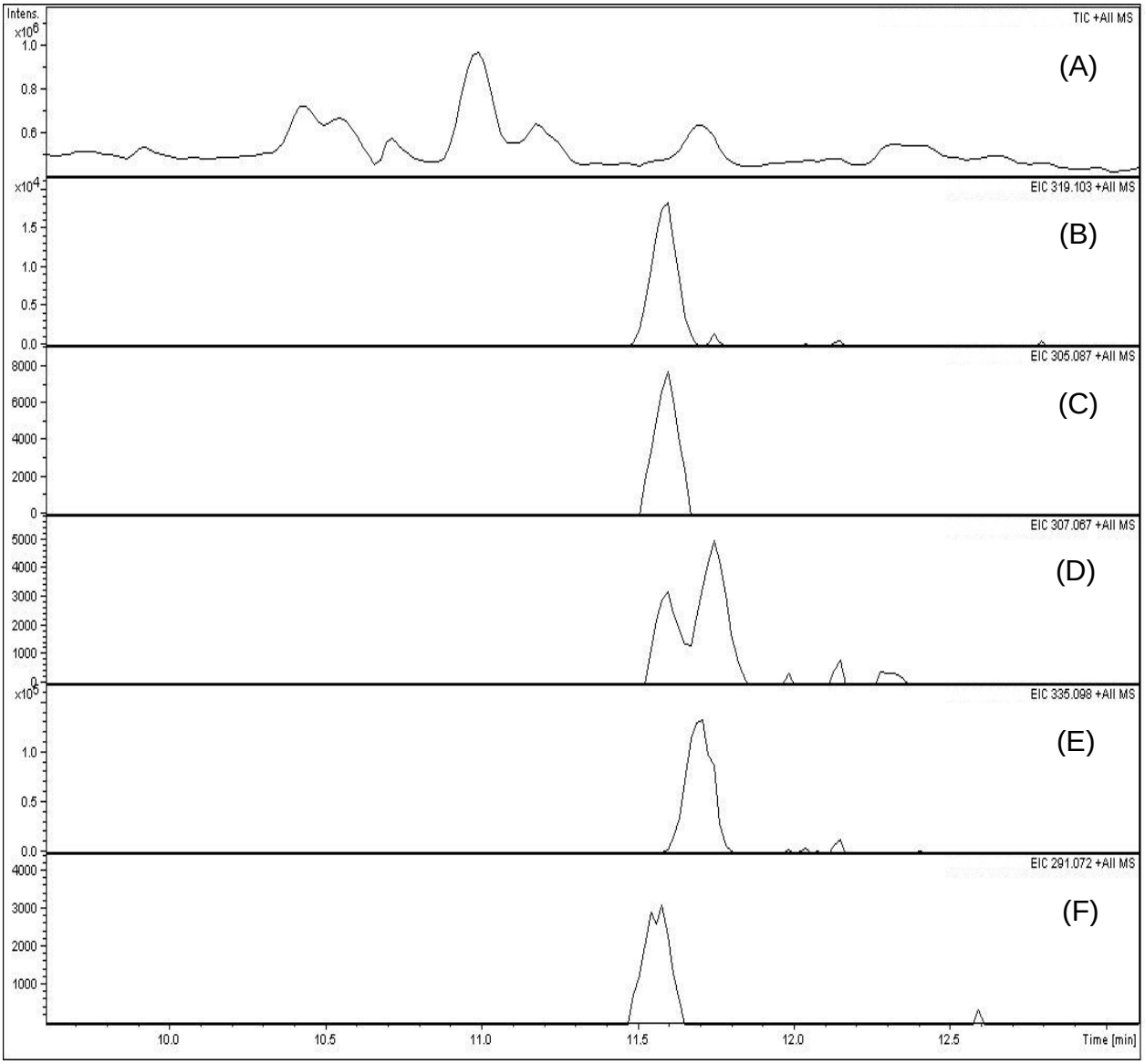


Figure 8. *Total ion (A) and mass electropherograms of the theoretical masses of chlorpromazine (CPZ) and metabolites in urine of a suicide case by hanging: m/z 319.103 (B, CPZ); m/z 305.087 (C, desmethyl-CPZ); m/z 307.067 (D, didesmethyl-CPZ sulfoxide; the less intense peak on the left is the isotopic M+2 peak of desmethyl-CPZ); m/z 335.098 (E, mass corresponding to 4 CPZ metabolites with identical chemical formula: CPZ N-oxide, CPZ sulphoxide, 7-hydroxy-CPZ; 8-Hydroxy-CPZ; the peak profile indicates at least 2 almost overlapping isobaric compounds); m/z 291.072 (F, didesmethyl-CPZ).*

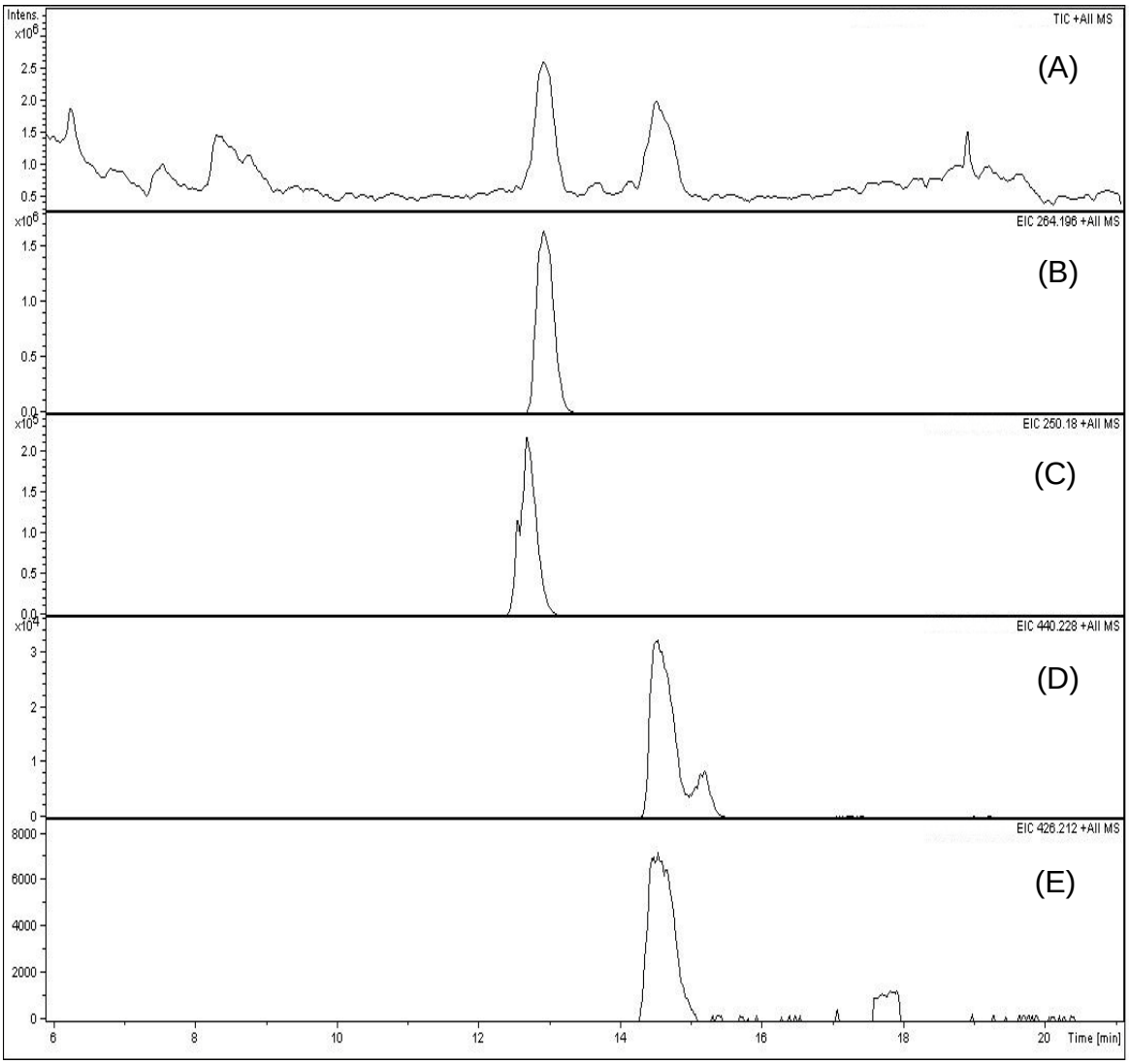


Figure 9. *Total ion (A) and mass electropherograms of the theoretical masses of tramadol (TMD) and metabolites in urine of a treated patient: m/z 264.196 (B, TMD); m/z 250.180 (C, mass corresponding to N- and O-desmethyl-TMD); m/z 440.228 (D, TMD glucuronide, not present in the database; it is worth noting that the 2 peaks are likely due to the formation of 2 diastereomeric glucuronides from racemic tramadol); m/z 426.212 (E, mass corresponding to N- and O-desmethyl-TMD glucuronide, both not present in the database).*

6.3. CONCLUSIONS

Our results demonstrate that Pubchem Compound PTRC subset database allows a much broader screening in biosamples by means of accurate mass/IP matching than other databases previously proposed. Despite minor “holes” and redundancies (which can be easily corrected), the database is extremely rich of parent drugs, metabolites, and related compounds. It can be easily downloaded free of charge from the internet, converted into a standalone database and implemented in any ESI/TOF-based screening method enabling what is, to our knowledge, the closest approach to the theoretical “general unknown search” of PTRC.

The database size (about 50.500 records) allows fast search (fractions of a second) and makes it usable in automated routines of repeated search for different measured masses. However, the disadvantage of comprehensiveness is a higher probability to retrieve different PTRC with identical CF than with smaller databases. Examples previously illustrated clearly show that with such a large database accurate mass/IP data do not always provide enough information to unequivocally identify the correct candidate.

Additional information must be used in order to focus search more tightly. Anamnestic/circumstantial data, and concomitant presence of parent compound and metabolites are very helpful but are not always available/reliable. However, it must be emphasized that, particularly when screening metabolites-rich matrices (*e.g.* urine), the presence of many metabolites in the database helps corroborating the identification of a parent compound. Useful information is provided by selective sample preparation: extraction/purification of basic/neutral compounds was applied in the examples illustrated here, thus allowing to exclude acidic compounds.

Retention or migration behaviour (depending on the separation applied before ESI-TOF: LC or CE), provide valued information for discriminating among isobaric compounds (Hill D.W. et al., 1994; Elliott S.P. et al., 1998; Boone C.M. et al., 1999). In this context, two approaches seem to be viable: (a) building-up a retention/migration database (Ojanpera S. et al., 2006) and (b) using information already available that may be correlated to retention/migration behaviour. Needless to say that approach (a) would limit the screening to compounds predicted by the analyst, not differently from what happens with in-house mass spectral or CF databases. Approach (b) does not require to build-up a database. Computational chemistry allows to estimate rather precisely some chemical/physical parameters starting from structure, *e.g.* octanol/water partition coefficient (P) or aqueous acid/base dissociation constant(s). For example, the correlation coefficient between estimated and experimental Log*P* of 1831 organic compounds was found to be 0.968 (with 0.37 SD) (Wang R. et al., 1997), and the accuracy of pKa calculations is typically better than ± 0.5 units (Rekker R. et al., 1993; Slater B. et al., 1994). The strength of correlation between these parameters and retention/migration behaviour (*i.e.* correlation between RP LC retention and calculated Log*P* and between electrophoretic mobility and calculated pKa) and its usefulness in discriminating different isobaric compounds will be matter of future research.

7. "METABOLOMIC" APPROACH

The combination of Atmospheric Pressure Ionization (API) with High resolution mass spectrometry (HRMS) detection has the potential for greatly expanding our capabilities in general unknown screening in different application fields (Fang L. et al., 2003; Ibanez M. et al., 2005; Ojanpera S. et al., 2006; Sancho J.V. et al., 2006; Kolmonen M. et al., 2007; Ferrer I. et al.,2007; Polettini A. et al., 2008). While API sources widen the range of identifiable compounds in terms of molecular weight, polarity, pK_a and chemical/thermal stability range and produce generally intense protonated/deprotonated molecular ions (Politi L. et al., 2006), HRMS enables the identification of the chemical formula of the protonated/deprotonated molecular ion through the accurate measure of its mass and isotopic pattern (Pelzing M. et al., 2004). Therefore, the API/HRMS combination permits a substantial step forward with respect to the so-called library search strategy. In fact, not only it avoids the build-up of expensive, time consuming and, consequently, limited mass spectral reference databases, but it also virtually eliminates the bias in identification due to the *a priori* selection of identifiable compounds (with the library-search strategy only compounds that the scientist decides to include in the database, and whose mass spectra may be acquired by analysing a pure standard or obtained through other means, have a chance of being identified) (Sauvage F.L. et al., 2006; Dresen S. et al., 2006).

However, leaving the old pathway for a new direction is not free from risks and uncertainties. As a matter of fact, the major limitation of identification based on the accurate determination of mass and isotopic pattern is that there is no biunique relation between chemical formula and structural formula (Maurer H.H. et al., 2007), the number of isomers being directly proportional to the extension of the reference

compounds database. As a result, by expanding the range of identifiable compounds (*i.e.* by enlarging the reference database of compounds) a longer list of candidates to identification for a given unknown compound is obtained. Recently, we have shown that the use of a database containing data on over 50.000 pharmaco-toxicologically relevant compounds (PTRC Database) and their phase I and II metabolites (Polettini A. et al., 2008) - a subset of PubChem Compounds (www.pubchem.ncbi.nlm.nih.gov, 2007) - allows a much broader screening with respect to databases containing a few hundreds of reference compounds, thus significantly reducing the bias in identification. However, in many cases the search in such a large database retrieves a list of 6-10 or even more candidate isomers.

Additional tools are therefore required in order to discriminate between compounds with the same elemental composition. In the attempt to achieve this goal, we have developed a novel "metabolomic" approach based on the detection of a predefined list of metabolites using high resolution mass spectrometry. The detection of common metabolites by means of extracting the ion chromatograms of the expected protonated molecular ions, using both low resolution mass spectrometry (in combination with fragmentation studies) (Gangl E. et al., 2002; Lafaye A. et al., 2003) and high resolution mass spectrometry (Plumb R.S. et al., 2003; Zhu M. et al., 2006; Zhang H. et al., 2009) is a well known strategy in drug metabolism studies. In the present work we have applied this strategy in the field of toxicological screening in order to discriminate between isomers.

The approach here proposed consists in the following steps: (a) the mass and isotopic pattern of the unknown compound are searched against the PTRC database and a list of candidates is retrieved; (b) starting from the mass of an unknown compound, mass shifts corresponding to 23 fixed biotransformations (*e.g.*

demethylation, hydroxylation, glucuronidation, etc.) are calculated and the corresponding mass chromatograms are extracted from the total ion current (TIC) in order to search for metabolite peaks; (c) for each compound retrieved in the list of candidates, the number of different functional groups in the molecule (N,O,S-methyls, hydroxyls, acetyls, etc.) is calculated using E-Dragon software (Talete srl, Milan, Italy) (Tetko I.V. et al., 2005); (d) the presence of metabolites in the TIC is matched with functional groups data in order to exclude candidates whose structure is not compatible with one or more of the observed biotransformations (*e.g.* loss of methyl from a structure not bearing methyls, glucuronidation on a structure not bearing any site susceptible to conjugation).

The performance of this "metabolomic" approach in shortening the list of candidate isomers was evaluated on PTRC detected in different biological samples (urine, hair, blood) and its advantages and limitations are illustrated in this paper through selected examples.

7.1. MATERIALS AND METHODS

7.1.1. CHEMICAL AND STANDARDS

All solvents and reagents were of HPLC or analytical grade and were purchased as described previously in the 5.1.1. section.

7.1.2. SAMPLE COLLECTION AND PREPARATION

Hair, blood and urine samples submitted to CE/MS analysis were treated as described in section 5.1.2. Urine samples analysed by LC/MS were first centrifuged

at 10.000 rpm (5590 x g) for 5 min and then diluted 1:20 with 0.1% formic acid before injection (Table 3).

Table 3. *Samples collected from real positive cases. The number, the separation system utilized for coupling with TOFMS, the reason of the toxicological analysis and the type of biological matrices are specified.*

N°	Separation system	Subject Initials	Circumstantial Data	Sample
1	CE	D.M.	Suspected Lethal Poisoning	Blood
2	CE	C.R.	Suspected Lethal Poisoning	Blood
3	CE	B.M.	Suspected Lethal Poisoning	Blood
4	CE	L.R.P.	Suspected Lethal Poisoning	Urine
5	CE	L.M.	Submitted to pharmaceutical substitution treatment	Urine
6	CE	B.T.	Admitted to emergency room for suspected intoxication	Urine
7	CE	S.A.	Suspected Lethal Poisoning	Urine
8	CE	C.R.	Relapse during substitution treatment	Hair
9	CE	A.D.	Relapse during substitution treatment	Hair
10	CE	L.A.	Relapse during substitution treatment	Hair
11	CE	Z.G.	Relapse during substitution treatment	Hair
12	CE	C.W.	Relapse during substitution treatment	Hair
13	CE	G.A.	Relapse during substitution treatment	Hair
14	CE	C.P.	Relapse during substitution treatment	Hair
15	CE	F.A.	Relapse during substitution treatment	Hair
16	CE	L.S.	Relapse during substitution treatment	Hair
17	CE	D.S.A.	Relapse during substitution treatment	Hair
18	CE	B.F.	Suspected Lethal Poisoning	Hair
19	LC	L.L.	Suspected Lethal Poisoning	Urine
20	LC	S.M.	Submitted to pharmaceutical substitution treatment	Urine
21	LC	S.S.	Suspected Suicidal Self-Poisoning	Urine
22	LC	O.C.	Suspected Lethal Poisoning	Urine
23	LC	F.A.	Suspected Lethal Poisoning	Urine
24	LC	V.M.	Admitted to emergency room for suspected intoxication	Urine
25	LC	P.G.	Suspected Lethal Poisoning	Urine
26	LC	M.G.	Suspected Lethal Poisoning	Urine
27	LC	C.A.	Suspected Lethal Poisoning	Urine
28	LC	R.R.	Suspected Lethal Poisoning	Urine
29	LC	P.S	Suspected Lethal Poisoning	Urine
30	LC	D.L.	Checked for fitness to drive	Urine
31	LC	E.L.	Submitted to pharmaceutical substitution treatment	Urine
32	LC	S.P.	Suspected Lethal Poisoning	Urine
33	LC	A.R.	Submitted to pharmaceutical substitution treatment	Urine
34	LC	S.L.	Checked for fitness to drive	Urine
35	LC	R.B.	Suspected Lethal Poisoning	Urine
36	LC	G.B.M.	Suspected Lethal Poisoning	Urine
37	LC	B.L.	Submitted to pharmaceutical substitution treatment	Urine

7.1.3. LC INSTRUMENTATION

An Agilent 1100 series liquid chromatograph connected with a TOF analyser interfaced with an electrospray source was used. The chromatograph was equipped with a ZORBAX-Eclipse-C18 Agilent column (2.1 x 150 mm i.d., 3.5 μm particle size), kept at 25 °C. The mobile phase consisted of 0.1% formic acid (A) and acetonitrile (B), the flow rate was 200 μl/min. The gradient consisted of 5% solvent B for 2 min, from 5% to 60% for 28 min, constant at 60% for 5 min, held at 90% for 5 min and followed by immediate return to the initial conditions (5% B) and 10 min re-equilibration. The injection volume was 8 μl. The instrument controls were performed with HyStar 3.1 and micrOTOFcontrol software (Bruker Daltonics).

7.1.4. CE/MS INSTRUMENTATION

All the apparatus and the experimental conditions applied were described in section 5.1.3.

7.1.5. DATA ANALYSIS

Data processing was carried out using Data Analysis software (Version 3.4, Bruker Daltonics), Microsoft Excel (2003 version), and E-Dragon (free internet 1.0 version, Milan, Italy), a software calculating several physical and chemical descriptors, including the type and number of functional groups of a given molecule, starting from different molecular file formats, including the Simplified Molecular Input Line Entry Specification (SMILES). Chromatographic/electrophoretic peaks pertaining to possible unknown compounds were searched within the row data by using background subtraction and peak deconvolution/mass spectral purification tools available in the Data Analysis Software. Then, the mass value of the base mass peak

was searched (±10 ppm tolerance), after proton subtraction, into a subset of the PubChem Compound database, named PTRC subset database, containing over 50.000 compounds of toxicological interest and many phase I and II metabolites (Polettini A. et al., 2008). For each chemical formula retrieved in the list of candidate compounds, mass error (difference between measured and theoretical mass), and sigma (a parameter, calculated by the Bruker software, accounting for the difference between theoretical and measured IP; the lower the value, the better the matching) were calculated. Structural isomers, salts, duplicates and compounds with sigma > 0.03 were excluded from the candidates list. For each remaining substance the number of a predefined set of functional groups (*e.g.* methyl, carboxyl, acyl groups) was calculated by using E-Dragon. Subsequently, a predefined set of 23 biotransformations (Table 4) from the mass of the unknown was searched and the corresponding extracted ion chromatograms/electropherograms (EICs/EIEs) were created with a 0.01 *m/z* mass width. The compatibility of E-Dragon data (Tables 5-10) for each candidate (number and type of functional groups) and metabolites detected (within a mass tolerance of ±10 ppm and a sigma tolerance of ≤ 0.03) was automatedly evaluated using Excel in order to exclude candidates with a molecular structure not compatible with one or more of the observed biotransformations.

The procedure was applied on 121 PTRC containing drugs of abuse (*e.g.* cocaine, opiates, MDMA), antidepressants (*e.g.* trazodone, fluoxetine, citalopram, venlafaxine, amitriptyline), benzodiazepines (*e.g.* flurazepam), antipsychotics (*e.g.* amisulpride), phenotiazines (*e.g.* promazine, pericyazine, chlorpromazine), histamine H2-receptor antagonists (*e.g.* ranitidine), anticonvulsants (*e.g.* gabapentin, carbamazepine), acetyl-cholinesterase inhibitors (*e.g.* rivastigmine), beta-blocker (*e.g.* bisoprolol), anti-retroviral agents (*e.g.* emtricitabine, tenofovir), antihistamine (*e.g.* cetirizine) and their

phase I metabolites. Specifically, 73 PTRC were detected in different biosamples (blood, urine and hair) subjected to CE/TOF screening, 48 compounds were detected in urine samples submitted to LC/TOF analysis.

Table 4. *Predefined set of 23 biotransformations with calculated mass shifts and the corresponding chemical formula modification.*

BIOTRANSFORMATION	MASS SHIFT	CHEMICAL FORMULA MODIFICATION
Demethylation	-14.0157	- CH2
Didemethylation, deethylation	-28.0313	- CH2CH2
Hydroxylation, N-Oxidation, Sulfoxide formation	15.9949	+ O
Hydration	18.0106	+ H2O
Deacetylation	-42.0106	- OC2H2
Acetylation	42.0106	+ OC2H2
Sulfonation	79.9568	+ SO3
Glucuronidation	176.0321	+ C6H8O6
Hydrogenation	2.0157	+ H2
Decarboxyilation	-43.9898	- O2C
Dehydrogenation	-2.0157	- H2
Methylation	14.0157	+ CH2
Nitro Reduction	-29.9742	- O2 + H2
Hydroxylation + Demethylation	1.9793	+ O - CH2
Hydrogenation + Demethylation	-12.0000	- C
Dehydrogenation + Demethylation	-16.0313	- CH4
Deamination	1.03160	- NH3 + O
Demethylation + Glucuronidation	162.0164	+ C5H6O6
Hydroxylation + Glucuronidation	192.0270	+ C6H8O7
Deacetylation + Glucuronidation	134.0215	+ C4H6O5
Oxidative desulfuration	-15.9772	- S + O
Conjugation with GSH	305.0682	+ C10H15N3O6S
Conjugation with glucose	162.0528	+ C6H10O5

7.2. RESULTS

The performance of the "metabolomic" approach in shortening the list of candidates was evaluated by analyzing real positive cases. Selected examples of the application of the procedure are described below.

Example 1 shows the capability of the method in discriminating among isomers of promazine (target compound) in the urine of an intoxication case. The analysis was performed by LC-MS after sample dilution in the initial mobile phase. Starting from the mass of the unknown compound (m/z 285.1415), a search in the PTRC subset database was performed and a list of 7 candidates and two chemical formulas within the given mass tolerance (±10 ppm) and sigma fit tolerance (≤ 0.03) was obtained (Figure 10 A). Then, mass shifts corresponding to the 23 predefined biotransformations were calculated and the corresponding mass chromatograms were extracted from the Total Ion Current (TIC) in order to search for metabolite peaks. One peak was detected in the demethylation chromatogram, 2 in the hydroxylation, 1 in the glucuronidation, 1 in the combined hydroxylation+demethylation and finally 1 peak in the combined hydroxylation+glucuronidation chromatogram (Figure 11). The matching of these acquired data with the functional groups data obtained with E-Dragon software for the 7 candidates permitted the exclusion of 4 of them (compounds 1, 4 and 5 do not have any site available for methylation, whereas in the case of compound 3 the direct glucuronidation in detectable amounts of a tertiary amino group attached to an aromatic ring was considered as unlikely). As a result, the only compounds susceptible to undergo the observed metabolisations were isomethazine, promethazine and promazine (Tables 5, 6).

Example 2 illustrates the performance of the "metabolomic" approach in the identification of venlafaxine and its metabolites in urine of a suicide case. Analyses were carried out by LC-MS after sample dilution in the initial mobile phase. The search of the mass m/z 278.2128 in the PTRC database retrieved a list of 6 candidates. Two peaks were detected in the demethylation chromatogram and 1 in

the glucuronidation chromatogram (Table 7). The matching with the functional groups data for the 6 candidates (Table 8) allowed to exclude 4 of them. In fact, the only compounds bearing at least 2 sites for demethylation (*i.e.* methyls attached to an heteroatom) and 1 for glucuronidation (*i.e.* either hydroxyls, carboxyls, sulphidryls and amino groups) were rds-127 and venlafaxine.

Example 3 concerns monodesmethylcitalopram, primary metabolite of citalopram, in the hair sample of a drug abuser under therapeutic treatment examined by CE-MS. The unknown compound exhibited a base mass peak at 311.1551 m/z. After the search in the PTRC subset database and the application of the sigma filter, a list of 4 candidates and 3 chemical formulas were retrieved. By extracting the EIEs for the 23 predefined biotransformations, the following metabolites were detected: 1 peak for demethylation, 1 peak for hydroxylation/N-oxidation/sulfoxide formation (resulting in all cases in the addition of one O atom), and 1 for methylation (Table 9). Again, starting from the SMILES, the number of different functional groups in the molecules of each candidate in the retrieved list was calculated using E-Dragon software (Table 10). Data processing showed that "monodesmethylcitalopram" was the only structure compatible with the observed metabolisations. In fact, 9-Daeta and Bifonazole do not have any methyl available for demethylation, whereas meprazine does not have any site available for methylation (Figure 10 B).

Example 4 refers to the case of methadone in the urine of a patient under treatment (data not shown). Analyses were carried out by LC-MS. From the mass of the unknown (310.2170 m/z), a list of 10 candidates with identical chemical formula was retrieved. However, in this case, no metabolites among those included in the predefined list were detected. As a result, the "metabolomic" approach was not helpful in reducing the list of candidates.

In order to evaluate the performance of the method in reducing the length of the list of candidates on a sufficiently large number of cases, 121 PTRC including parent compounds and their phase I metabolites in several biosamples (urine, blood, hair) were processed as described in the examples shown earlier (Table 11). The mean list length (MLL) of compounds was 6.71 ± 4.66 (median 6, range 1-28) before the application of the "metabolomic" approach and after the application it was shortened to 3.94 ± 3.07 (median 3, range 1-17). By grouping PTRC cases by separation technique (either CE or LC) and the reductions in the MLL were obtained with the application of the "metabolomic" approach are the following: from 6.85 ± 5.10 (median 7, range 1-28) to 4.16 ± 3.11 (median 3, range 1-13) for PTRC and phase I metabolites analysed by CE/TOF (73 compounds); from 6.50 ± 3.94 (median 7, range 2-21) to 3.60 ± 3.00 (median 3, range 1-17) for compounds analysed by LC/TOF (48 compounds).

Further evaluations were carried out in order to evaluate the performance of the procedure in shortening the MLL for compounds detected in different matrices (Table 11). For PTRC identified in urine samples (n = 69) the reduction in the MLL was from 6.59 ± 4.29 (median 6, range 1-21) to 3.75 ± 3.20 (median 3, range 1-17), whereas in the case of PTRC detected in hair (n = 44) the MLL was reduced from 7.41 ± 5.55 (median 7, range 1-28) to 4.77 ± 3.40 (median 3.50, range 1-13). For the 9 PTRC identified in blood the MLL was shortened from 5.22 ± 2.86 (median 6, range 1-8) to 2.67 ± 2.06 (median 2, range 1-7).

Table 5. *Number of peaks detected for each of the mass chromatograms corresponding to the 23 predefined biotransformations calculated from 285.1415 m/z (example 1).*

Biotransformation	**m/z**	**Number of peaks detected**	**Chemical Formula**	**Error (ppm)**	**σ**
Demethylation	**271.1258**	**1**	**C16H19N2S**	**3.3**	**0.0253**
Didemethylation, deethylation	257.1102	0			
Hydroxylation, N-Oxidation, Sulfoxide formation	**301.1364**	**2**	**C17H21N2OS**	**- 2.0**	**0.0234**
Hydration	303.1521	0			
Deacetylation	243.1309	0			
Acetylation	327.1521	0			
Sulfonation	365.0983	0			
Glucuronidation	**461.1736**	**1**	**C23H29N2O6S**	**1.8**	**0.0239**
Hydrogenation	287.1572	0			
Decarboxyilation	241.1517	0			
Dehydrogenation	283.1258	0			
Methylation	299.1572	0			
Nitro Reduction	255.1673	0			
Hydroxylation + Demethylation	**287.1208**	**1**	**C16H19N2OS**	**- 1.9**	**0.0131**
Hydrogenation + Demethylation	273.1415	0			
Dehydrogenation + Demethylation	269.1102	0			
Deamination	286.1731	0			
Demethylation + Glucuronidation	447.1579	0			
Hydroxylation + Glucuronidation	**477.1685**	**1**	**C23H29N2O7S**	**- 2.0**	**0.0093**
Deacetylation + Glucuronidation	419.1630	0			
Oxidative desulfuration	269.1643	0			
conjugation with GSH	590.2097	0			
conjugation with glucose	447.1943	0			

Table 6. *Candidates list obtained from the PTRC database through the search of mass and isotopic pattern (considered parameters: exact mass within ±10 ppm from the mass of the unknown; sigma≤ 0.03), type and number of functional groups and evaluation of compatibility with the observed biotransformations (example 1).*

No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Compound Name	**6-Bdbda**	**Isomethazine**	**Michler's thione**	**9-(p-Toluidino) acridine**	**1,1'-Bi(2-naphthalenylamine)**	**Promethazine**	**Promazine**
Chemical Formula	**C20H16N2**	**C17H20N2S**	**C17H20N2S**	**C20H16N2**	**C20H16N2**	**C17H20N2S**	**C17H20N2S**
CID PubChem	**133478**	**110669**	**71045**	**51971**	**20571**	**4927**	**4926**
Functional Groups							
Methyl groups (demethylation, didemethylation)	0	2	4	0	0	2	2
Ethyl groups (deethylation)	0	1	0	0	0	1	2
Alcohol, phenol, enol, carboxyl OH (glucuronidation)	0	0	0	0	0	0	0
Other H attached to heteroatom, es. amino groups, including aliphatic tertiary amino groups (glucuronidation)	1	1	0	1	4	1	1
H attached to heteroatom (methylation, acetylation)	1	0	0	1	4	0	0
Carboxyl groups (decarboxylation)	0	0	0	0	0	0	0
Nitro groups (nitro reduction)	0	0	0	0	0	0	0
Oxiranes (hydration)	0	0	0	0	0	0	0
H attached to heteroatom – carboxyls	1	0	0	1	4	0	0
Primary amino group (deamination)	0	0	0	0	2	0	0
Acyl groups (Deacylation, deacetylation)	0	0	0	0	0	0	0
Sulphurs (oxidative desulfuration)	0	1	1	0	0	1	1
Number of incompatibilities	2	0	1	2	2	0	0
COMPATIBLE	NO	**YES**	NO	NO	NO	**YES**	**YES**

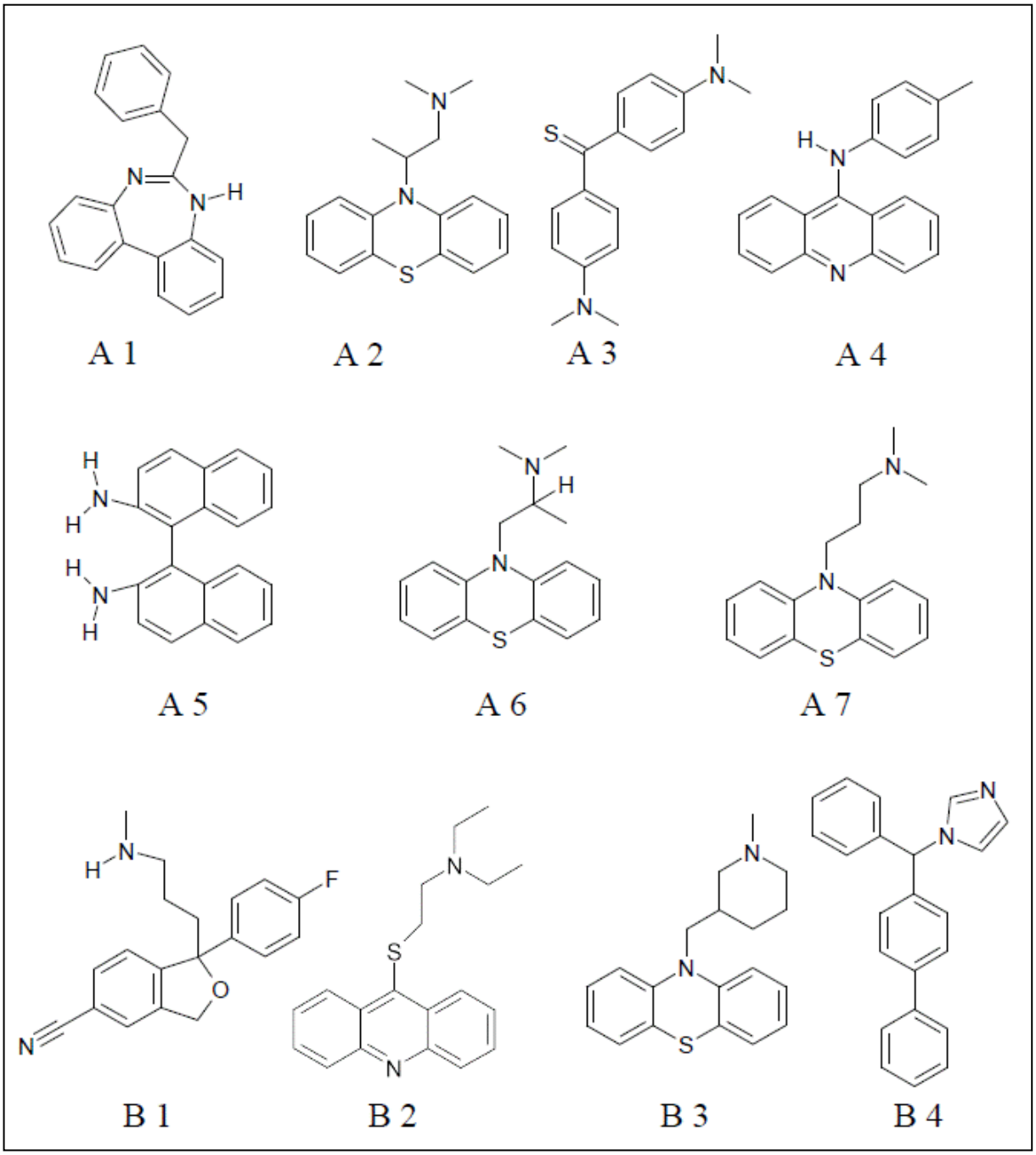


Figure 10. *Structures of the compounds in the candidates list for example 1 [(A) 1, 6-Bdbda; 2, Isomethazine; 3, Michler's thione; 4, 9-(p-Toluidino) acridine; 5, 1,1'-Bi(2-naphthalenylamine); 6, Promethazine; and 7, Promazine] and example 3 [(B) 1, monodesmethylcitalopram; 2, 9-daeta; 3, meprazine; 4, bifonazole].*

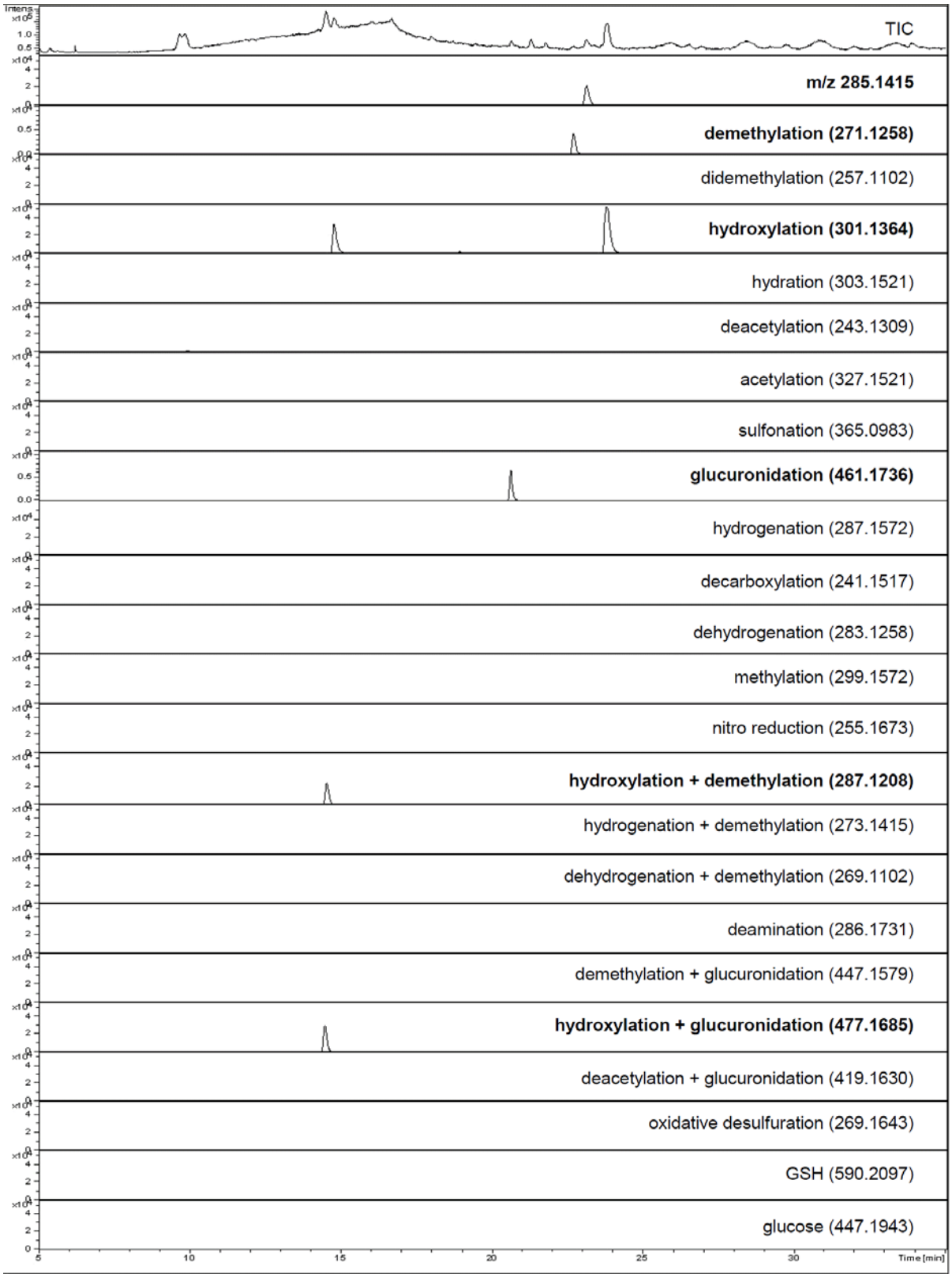


Figure 11. *Total ion and mass chromatograms corresponding to the 23 biotransformations calculated from the 285.1415 m/z (example 1).*

Table 7. *Number of peaks detected for each of the mass chromatograms corresponding to the 23 predefined biotransformations calculated from 278.2128 m/z (example 2).*

Biotransformation	**m/z**	**Number of peaks detected**	**Chemical Formula**	**Error (ppm)**	**σ**
Demethylation	**264.1971**	**2**	**C16H26NO2**	**-6.4**	**0.0220**
Didemethylation, deethylation	250.1815	0			
Hydroxylation, N-Oxidation, Sulfoxide formation	294.2077	0			
Hydration	296.2234	0			
Deacetylation	236.2022	0			
Acetylation	320.2234	0			
Sulfonation	358.1696	0			
Glucuronidation	**454.2449**	**1**	**C23H36NO8**	**6.2**	**0.0282**
Hydrogenation	280.2285	0			
Decarboxyilation	234.2230	0			
Dehydrogenation	276.1971	0			
Methylation	292.2285	0			
Nitro Reduction	248.2386	0			
Hydroxylation + Demethylation	280.1921	0			
Hydrogenation + Demethylation	266.2128	0			
Dehydrogenation + Demethylation	262.1815	0			
Deamination	279.2444	0			
Demethylation + Glucuronidation	440.2292	0			
Hydroxylation + Glucuronidation	470.2398	0			
Deacetylation + Glucuronidation	412.2343	0			
Oxidative desulfuration	262.2356	0			
conjugation with GSH	583.2810	0			
conjugation with glucose	440.2656	0			

Table 8. *Candidates list obtained from the PTRC database through the search of mass and isotopic pattern (considered parameters: exact mass within ±10 ppm from the mass of the unknown; sigma≤ 0.03), type and number of functional groups and evaluation of compatibility with the observed biotransformations (example 2).*

No.	**1**	**2**	**3**	**4**	**5**	**6**
Compound Name	**Nchembio801-comp6**	**Rds-127**	**Padimite A**	**Padimate O**	**Azak**	**Venlafaxine**
Chemical Formula	**C17H27O2N**	**C17H27NO2**	**C17H27O2N**	**C17H27O2N**	**C17H27O2N**	**C17H27NO2**
CID PubChem	**5484725**	**134064**	**42851**	**30541**	**15967**	**5656**
Functional Groups						
Methyl groups (demethylation, didemethylation)	0	2	2	2	1	3
Ethyl groups (deethylation)	1	2	1	1	0	1
Alcohol, phenol, enol, carboxyl OH (glucuronidation)	1	0	0	0	0	1
Other H attached to heteroatom, es. amino groups, including tertiary aliphatic amino groups (glucuronidation)	1	1	0	0	1	1
H attached to heteroatom (methylation, acetylation)	2	0	0	0	1	1
Carboxyl groups (decarboxylation)	0	0	0	0	0	0
Nitro groups (nitro reduction)	0	0	0	0	0	0
Oxiranes (hydration)	0	0	0	0	0	0
H attached to heteroatom–carboxyls	2	0	0	0	1	1
Primary amino group (deamination)	0	0	0	0	0	0
Acyl groups (Deacylation, deacetylation)	0	0	1	1	0	0
Sulphurs (oxidative desulfuration)	0	0	0	0	0	0
Number of incompatibilities	2	0	1	1	1	0
COMPATIBLE	NO	**YES**	NO	NO	NO	**YES**

Table 9. *Number of peaks detected for each of the mass chromatograms corresponding to the 23 predefined biotransformations calculated from 311.1551 m/z (example 3).*

Biotransformation	**m/z**	**Number of peaks detected**	**Chemical Formula**	**Error (ppm)**	**σ**
Demethylation	**297.1394**	**1**	**C18H18FN2O**	**1.9**	**0.0131**
Didemethylation, deethylation	283.1238	0			
Hydroxylation, N-Oxidation, Sulfoxide formation	**327.1500**	**1**	**C19H20FN2O2**	**- 2.1**	**0.0280**
Hydration	329.1657	0			
Deacetylation	269.1445	0			
Acetylation	353.1657	0			
Sulfonation	391.1119	0			
Glucuronidation	487.1872	0			
Hydrogenation	313.1708	0			
Decarboxyilation	267.1653	0			
Dehydrogenation	*309.1394*	*1*	*C19H18FN2O*	*3.0*	*0.5650 **
Methylation	**325.1708**	**1**	**C20H22FN2O**	**6.0**	**0.0150**
Nitro Reduction	281.1809	0			
Hydroxylation + Demethylation	313.1344	0			
Hydrogenation + Demethylation	299.1551	0			
Dehydrogenation + Demethylation	295.1238	0			
Deamination	312.1867	0			
Demethylation + Glucuronidation	473.1715	0			
Hydroxylation + Glucuronidation	503.1821	0			
Deacetylation + Glucuronidation	445.1766	0			
Oxidative desulfuration	295.1779	0			
conjugation with GSH	616.2233	0			
conjugation with glucose	473.2079	0			

**Not considered for data processing (sigma ≥ 0.03).*

Table 10. *Candidates list obtained from the PTRC database through the search of mass and isotopic pattern (considered parameters: exact mass within ±10 ppm from the mass of the unknown; sigma≤ 0.03), type and number of functional groups and evaluation of compatibility with the observed biotransformations (example 3).*

No.	**1**	**2**	**3**	**4**
Compound Name	**Monodesmethylcitalopram**	**9-Daeta**	**Meprazine**	**Bifonazole**
Chemical Formula	**C19H19N2OF**	**C19H22N2S**	**C19H22N2S**	**C22H18N2**
CID	**162180**	**124077**	**6075**	**2378**
Functional Groups				
Methyl groups (demethylation, didemethylation)	1	0	1	0
Ethyl groups (deethylation)	2	4	3	0
Alcohol, phenol, enol, carboxyl OH (glucuronidation)	0	0	0	0
Other H attached to heteroatom, es. amino groups, including tertiary aliphatic amino groups (glucuronidation)	1	0	0	0
H attached to heteroatom (methylation, acetylation)	1	0	0	0
Carboxyl groups (decarboxylation)	0	0	0	0
Nitro groups (nitro reduction)	0	0	0	0
Oxiranes (hydration)	0	0	0	0
H attached to heteroatom - carboxyls	1	0	0	0
Primary amino group (deamination)	0	0	0	0
Acyl groups (Deacylation, deacetylation)	0	0	0	0
Sulphurs (oxidative desulfuration)	0	1	1	0
Number of incompatibilities	0	2	1	2
COMPATIBLE	**YES**	NO	NO	NO

Table 11. *Before and after the application of the "metabolomic" approach, mean, standard deviation, median, minimum and maximum number of compounds in the candidates list were calculated for each group of compounds.*

Sample	**Separation**	**Number of PTRC**	**"Metabolomic" Processing**	**Mean ± St.Dev.**	**Median (range)**
urine, blood, hair	LC or CE	121	NO	6.71±4.66	6 (1-28)
			YES	3.94±3.07	3 (1-17)
urine, blood, hair	CE	73	NO	6.85±5.10	7 (1-28)
			YES	4.16±3.11	3 (1-13)
urine	LC	48	NO	6.50±3.94	7 (2-21)
			YES	3.60±3.00	3 (1-17)
urine	LC or CE	69	NO	6.59±4.29	6 (1-21)
			YES	3.75±3.20	3 (1-17)
blood	CE	9	NO	5.22±2.86	6 (1-8)
			YES	2.67±2.06	2 (1-7)
hair	CE	44	NO	7.41±5.55	7 (1-28)
			YES	4.77±3.40	3.5 (1.13)

7.3. DISCUSSION

High resolution mass spectrometry is a very attractive technique for general unknown screening applications as it allows to identify the chemical formula of an unknown compound with no need to build-up expensive, time-consuming, and limited databases of mass spectral and retention/migration reference data. The mass (±10 ppm) and sigma fit (<0.03) tolerances, defined based on our personal experience as well as on other authors' results (Ojanpera S. et al., 2006), proved to be effective in reducing the number of compatible chemical formulas to a minimum and broad enough so as to avoid the exclusion of the correct formula.

However, the relationship between chemical formula and molecular structure is not biunique and tools are necessary in order to identify the correct candidate within a list of isomers. The "metabolomic" approach here described proved to be effective in achieving this goal by reducing the MLL of candidates almost to the half of the initial value. Examples 1 and 2 are a clear demonstration of the fitness for the purpose of the described approach. In fact, in these case the MLL was shortened to more than half permitting to focus on the identification of the unknown (promazine and venlafaxine, respectively).

Example 3 is even more striking as in this case the MLL was reduced from 4 to only one candidate (desmethylcitalopram, *i.e.* the correct one). It must be stressed that in this example the peak detected in the methylation EIE actually corresponds to the parent compound citalopram, and not to a methylated metabolite. Nevertheless, this information was helpful in establishing that the unknown compound had at least one site on the molecule available for methylation and therefore all candidates not possessing any sites available for methylation could be excluded.

On the other hand, example 4 (methadone) represents the worse case possible, *i.e.* a situation in which the "metabolomic" approach does not provide any means to reduce the length of the list of candidates. As a matter of fact, methadone undergoes a rather atypical metabolic pattern leading to 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine (EDDP, [$-CH_4O$]) and to the further demethylated compound EMDP [$-C_2H_6O$] as main metabolites and, to a minor extent, to methadol [$+H_2$] and normethadol [$+H_2-CH_2$]. Among these metabolites only methadol and normethadol could have been detected by extracting the EIEs for the 23 predefined biotransformations. However, the hydrogenation and the hydrogenation+demethylation chromatograms were negative as were all the other chromatograms for the predefined metabolisations. Owing to these reasons none of the 10 candidates could be excluded from the list. It is important to note that, when known, specific metabolites (*i.e.* not included in the predefined list) for one or more of the suspected candidates can be searched for *a posteriori* in the raw data and, if present, they may help in discriminating the correct candidate. In example 4 the presence of peaks in the EICs of EDDP and EMDP (extracted after acknowledging the presence of methadone in the list of candidates) supported the identification of methadone.

By definition, in all cases were none of the predefined biotransformations are observed (*i.e.* no peaks are detected in the EICs/EIEs corresponding to the predefined metabolisations) the "metabolomic" approach is useless (unless specific metabolites for one or more candidate are known, as in the case of methadone). In fact, we decided to avoid using the absence of a metabolite as a relevant information for data processing, for the simple reason that the absence of a metabolite does not imply that its formation is not possible given the molecular structure of the parent

compound (for example, one good reason for the absence of one metabolite is that the organism had no time enough to produce it in detectable amounts).

Another necessary consequence of the approach here described is that it should perform better with metabolites-rich matrices (*e.g.* urine) and for compounds that are far from the end of the metabolic pathway(s), *i.e.* that are more likely to produce a large number of metabolites (*e.g.* parent compound instead of a phase II metabolite). However, we obtained fairly good results in reducing the MLL also in PTRCs analysed in hair which typically contains less metabolites than urine. In addition, the inclusion in the predefined list of biotransformations of some bi-directional metabolic reactions (*e.g.* demethylation/methylation, deacetylation/acetylation, dehydrogenation/hydrogenation) allow the approach to work also for metabolites (see example 3 on desmethylcitalopram).

7.4. CONCLUSIONS AND FUTURE PERSPECTIVES

The combination of a large PTRC database of compound names, chemical formulas and exact masses (Kolmonen M. et al., 2007) with the "metabolomic" approach here described provides a powerful tool for toxicological screening purposes extending the number of identifiable compounds from a few hundreds (*e.g.* with the library search approach (Sauvage F.L. et al., 2006; Dresen S. et al., 2006) or the use of in-house chemical formula databases [Pelander A. et al., 2009]) to more than 50.000. Obviously, by extending the number of reference compounds in the database, a lower number of false negatives as well as a higher number of false positives (*i.e.* isomers of the correct candidate) are obtained. Therefore, tools to shorten the mean

list length of candidates are necessary. The "metabolomic" approach proved to serve for the purpose.

Although metabolomics has been already introduced recently in the field of general unknown screening (Pelander A. et al., 2009), its use has been limited to the prediction and identification of metabolites of a known compound in order to include their analytical data (exact monoisotopic masses and retention times) in a in-house reference database of compounds and metabolites of pharmaco/toxicological interest which contains, in the latest version available, 830 compounds. This is, to our knowledge, the first time that metabolites detection is exploited as a tool for discriminating among isomers in toxicological screening applications.

Our efforts are currently directed towards the full automation and interfacing of the procedure with the Bruker Data Analysis Software. Another aspect that we are investigating is the possibility to add a further discrimination tool to be implemented in the data processing of LC-TOF data. The strategy is based on the use of retention behaviour data in RPLC mode and their correlation with chemical/physical parameters (*e.g.* the octanol/water partition coefficient, *P*) estimated using computational chemistry from the molecular structure of a compound.

8. CHEMOMETRIC APPROACH

In the field of drugs metabolism studies, the application of low resolution mass spectrometry, combined with studies on fragmentation, on one side and of HRMS for the identification of metabolites through the extraction of the ion chromatograms of their anticipated protonated molecular ions on the other are the most common strategies (Zhu M., 2006; Zhang H. et al., 2008; Zhang H., 2009; Mortishire-Smith R.J. et al., 2009). On the other hand, the study of the molecular properties (*e.g.* lipophilicity) of a substance for the prediction of its chromatographic behaviour is of great interest for theoretical chemistry purposes (Remko M. et al., 2006; Etxebarria N. et al., 2008) and is well known as a strategy for compounds identification (Rourick R.A. et al., 1996; Tetko I.V. et al., 2001; Koh H.L. et al., 2003; Deineka V.I., 2006; D'Archivio A.A. et al., 2007; Lei B. et al., 2009). Different research groups have developed models for the prediction of retention time in a LC system by characterizing the chemical-physical features of a given compound (Loukas Y.L., 2000; Baczek T. et al., 2005; Put R. et al., 2007).

In the present work, we have implemented a chemometric strategy in the field of toxicological screening with the aim to discriminate between isobaric compounds in order to reduce the length of the list of candidates.

Our approach consists in developing and applying a model for roughly predicting RPLC retention behaviour of any candidate in the list starting from a number of molecular descriptors calculated from the representation of its structure with the Simplified Molecular Input Line Entry Specification (SMILES), using different

softwares (*i.e.* E-Dragon and ALOGPS) (http://www.vcclab.org, 2005) or obtained from the PubChem Compound Database. *The retention time range (expressed as RRT to Nalorphine as internal standard) estimated for each candidate is then compared with the measured RRT of the unknown. All candidates whose estimated retention behaviour is not compatible with the observed RRT are excluded from the list.*

The performance of the described approach, as well of its combination with the metabolomic approach previously illustrated, in reducing the list of isobaric compounds was evaluated through the analysis of real positive samples (urine, hair, blood).

8.1. MATERIALS AND METHODS

8.1.1. CHEMICAL AND STANDARDS

Reference standard of nalorphine [used as the internal standard (IS)], was purchased from S.A.L.A.R.S. (Como, Italy) and the stock solution was prepared in methanol at a concentration of 2 mg/ml. Water was purified using an aqua MAX-Ultra 370 Series water purification system (Young Lin Instrument, Anyang, Korea).

All solvents and reagents were of HPLC or analytical grade and were obtained as described before in section 5.1.1.

8.1.2. SAMPLE COLLECTION AND PREPARATION

Urine samples were treated as described in section 6.1.2.

8.1.3. LC/MS INSTRUMENTATION

An Agilent 1100 (Agilent Technologies) series LC system was used for chromatographic separation and was connected with a TOF analyser interfaced with an electrospray source. Instrumental features, parameters and conditions for the analysis were as described in section 6.1.3.

8.1.4. MATHEMATICAL MODEL FOR THE ESTIMATION OF RETENTION BEHAVIOUR

A mathematical model was developed in order to establish a correlation between the RPLC retention behavior of over 300 compounds of toxicological interests (Hill D.W. et al., 1994) with a group of molecular descriptors (including Log*P*, Hydrogen Bonds Donor Counts, No. of Nitrogen and Oxygen in the molecule, etc.) as calculated using E-Dragon, ALOGPS (http://www.vcclab.org, 2005) or obtained from Pubchem Compounds.

The model developed allows to explain as much as 88% of the scatter in retention times of this large group of compounds and therefore enables a rough estimation of the RPLC retention behaviour, expressed as an RRT range (Figure 12). In order to take into account the approximation of the RRT estimate a tolerance RRT range around the estimated value was adopted so as to avoid the exclusion of the right candidate from the list (see Figure 12).

8.1.5. DATA ANALYSIS

Data were submitted to the Bruker Data Analysis Processing and then subjected to the chemometric data processing, based on the evaluation of the compatibility between the measured RRT and that estimated from the molecular features of each

candidate (expressed as a range of RRT). In particular, the RRT range estimated for each candidate by applying the mathematical model described in the previous section was compared with the esperimental RRT of the unknown. Subsequently, all the compounds with an estimated RRT range not matching with that of the unknown were excluded from the list of candidates. Data processing was carried out on 43 substances of pharmaco-toxicological interest detected in urine samples, including drugs of abuse (*e.g.* cocaine, opiates), anticonvulsants (*e.g.*, gabapentin), antidepressants (*e.g.* citalopram, amitriptyline, venlafaxine), phenothiazines (*e.g.* promazine, levomepromazine, pericyazine), histamine H2-receptor antagonists (*e.g.* ranitidine), narcotic-analgesic (*e.g.* tramadol), beta-blocker (*e.g.* bisoprolol), anti-retroviral agents (*e.g.* emtricitabine, tenofovir), antihistamine (*e.g.* cetirizine) and their phase I metabolites, for a total analysis of the chemical-physical properties of 263 possible candidate compounds.

A parallel processing of the same raw data using the previously illustrated metabolomic approach was also carried out in order to compare the two approaches as well as their combination.

8.2. RESULTS

The performance of the chemometric approach into shortening the candidates list was evaluated through the application of the procedure to a group of real positive urine matrices submitted to the LC/TOF screening.

In order to illustrate the chemometric data processing procedure, selected examples are described below.

Example 1 shows the application of the procedure to the detection of **bisoprolol**. The urine sample was simply diluted in the initial mobile phase (1:20 v/v). Starting from the mass of the unknown compound (326.2345 m/z), the search against the PTRC subset database and the subsequent application of the sigma filter provided a list of 2 possible compounds: bisoprolol and dapiprazole. The detected peak had a RRT of 4.2. Then, the chromatographic retention obtained was matched with the expected retention range, allowing the exclusion of the dapiprazole from the list (RRT range of 0-7.01 for bisoprolol against 5.28-18.15 for dapiprazole. The only compound fitting the relation between retention estimated from the molecular properties and retention measured was bisoprolol (Table 12).

Example 2 concerns the application to the identification of **tenofovir** in urine of a HIV positive subject. The search of the mass 288.0881 m/z in the PTRC database retrieved a list of 5 possible candidates. A RRT of 0.49 was estimated for the unknown compound. The further evaluation of the chemical-physical properties for each compound has highlighted several differences (*e.g.* in the Log*P*). The matching of the chromatographic acquired data with estimated data from the analysis of the molecular properties permitted the exclusion of two substances out of five from the candidates list. In fact, both mecpac and 3-O-methyltolcapone presented an expected retention interval not including the RRT value experimentally observed. As a result they were eliminated and a list containing tenofovir, apropovir and 5-ddfa was finally retrieved (Table 12).

Example 3 illustrates the capability of the chemometric approach in reducing the length of the candidates list for **didesmethylcitalopram**, a metabolite of citalopram. From the mass 297.1420 m/z of the unknown compound, the search in the PTRC subset database produced a list of 6 candidates. Again, the RRT of the unknown

(4.64) was compared with the expected retention ranges calculated for all the candidates. As a result, 2 compounds were excluded: lophine and fluproquazone (Table 12).

Example 4 refers to the particular case of **methadone** in the urine of a patient under pharmaceutical treatment. Starting from the mass of the unknown (297.1420 m/z), through the use of the PTRC database, a list of 10 probable candidates was provided. Nevertheless, cause of almost all the compounds presented similar molecular properties, only 1 compound (1,6-bpep) was excluded from the list and 9 compounds remained (Table 12).

The performance of the chemometric approach was tested on 43 PTRC. The analysis was performed both for parent compounds and their phase I metabolites. The mean list length (MLL) of compounds was 6.02 ± 3.49 (median 6, range 2-21) before the application of the chemometric approach and was shortened to 3.09 ± 2.03 (median 2, range 1-9) after.

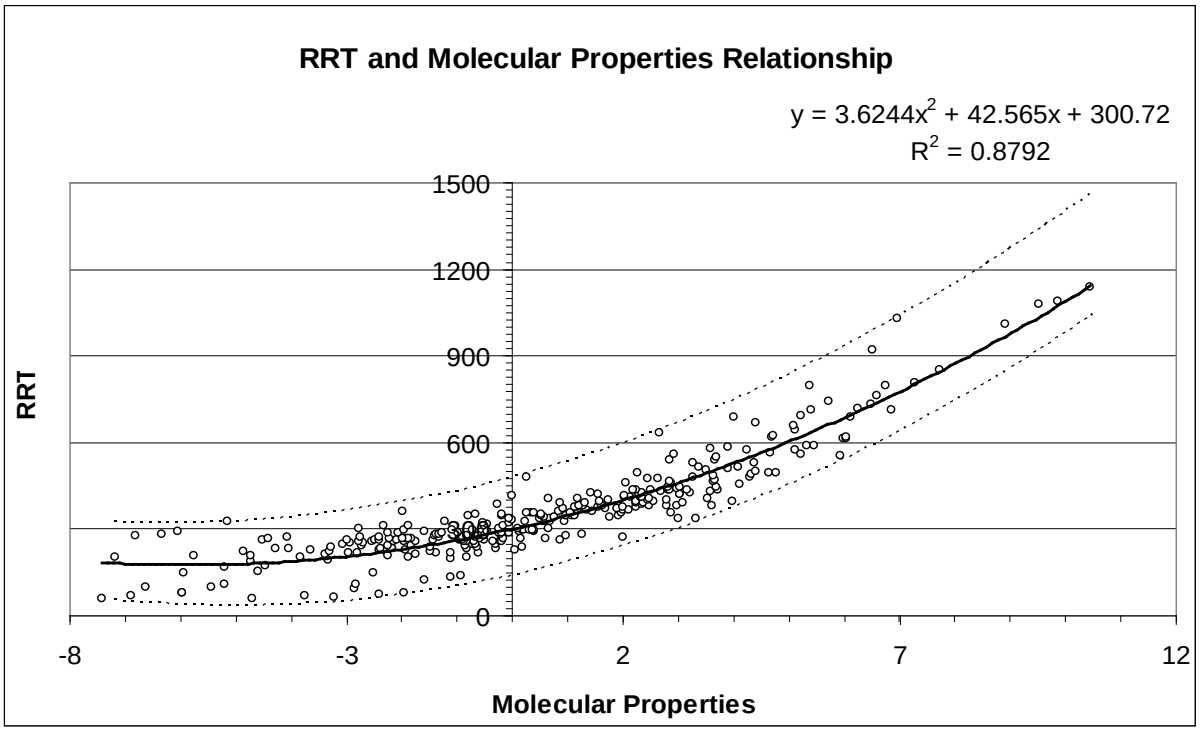


Figure 12. *Correlation curve between RRT (on the y axis) and molecular properties (on the x axis).*

Table 12. *For each given example, number and name of candidate, expected retention time, RRT measured finally evaluation of the compatibility were showed.*

		EXPECTED RETENTION INTERVAL			
N°	**CANDIDATE**	**MIN**	**MAX**	**RRT measured**	**COMPATIBILITY**
1	**Bisoprolol**	0	7.01	4.20	YES
2	Dapiprazole	5.28	18.15	4.20	NO
1	**Tenofovir**	0	4.19	0.49	YES
2	Apropovir	0	4.19	0.49	YES
3	Mecpac	0.74	13.05	0.49	NO
4	3-O-Methyltolcapone	6.04	18.99	0.49	NO
5	5-Ddfa	0	5.28	0.49	YES
1	Lophine	8.05	21.21	4.64	NO
2	**Didesmethylcitalopram**	0	8.56	4.64	YES
3	Fluproquazone	6.74	19.77	4.64	NO
4	PM3B	0.21	12.43	4.64	YES
5	Tacaryl	0.17	12.39	4.64	YES
6	Pyrathiazine	1.18	13.55	4.64	YES
1	Pulo'upone	5.60	18.51	6.34	YES
2	1,6-Bpep	8.72	21.94	6.34	NO
3	Aspaminol	0	10.09	6.34	YES
4	2-methyl-1,1-diphenyl-3-(1-piperidyl)propan-1-ol	0	10.09	6.34	YES
5	LY 106737	5.59	12.25	6.34	YES
6	Isomethadone	0.48	12.75	6.34	YES
7	**Methadone**	1.38	13.78	6.34	YES
8	Cephadol	0	11.61	6.34	YES
9	Benproperine	3.84	16.55	6.34	YES
10	Nchembio873-comp67	1.10	13.46	6.34	YES

8.3. DISCUSSION

The chemometric approach represents a further useful strategy into reduce the candidates list for the target unknown compound determination. The performance of the procedure is well shown in the example 1, for which its application permitted the clear identification of the bisoprolol as unique possible candidate.

Examples 2 and 3 are also a good demonstration of the fitness of our approach for discrimination purposes among several compounds. In fact, in both cases the MLL was shortened to almost the half permitting to focus the search on a lesser number of possible candidates for the final identification of the unknown target (tenofovir and didesmethylcitalopram, respectively).

For the example 4, the case of methadone, a list of 10 candidates was obtained. Unfortunately, as it happened in the "metabolomic" analysis, the chemometric approach did not provide a substantial reduction of the candidates list. In fact, cause of the similarity in the chemical-physical features among candidates, the match between the estimated and calculated RRT enabled the elimination of only 1 compound.

Of course, in all cases where the candidates in ~~of~~ the list present high likeness in the chemical/physical features, the discrimination power of the chemometric approach will be weaker. In addition, the chemometric approach has to be adjusted to the specific separation conditions adopted, whereas the "metabolomic" approach is universally applicable. On the other hand, an advantage of the chemometric approach over the metabolomic one is that it does not require metabolites to be present in the matrix.

8.4. CONCLUSIONS

As well known, the retention behaviour provides a relevant information for the discrimination among isobaric compounds. However, either measuring retention behaviour of isobaric candidates or predicting retention through the experimentally measured molecular properties requires the availability of standards and the build up of a database.

The chemometric approach that we have developed explores a new strategy, that is the estimation of the retention behaviour by using different molecular chemical/physical properties of a compound as calculated using a mathematical model. In this context, only the molecular structure of each candidate has to be known. Our experience demonstrated the suitability of the correlation between calculated chemical-physical properties of a given molecule and estimated retention time for general screening purposes. In fact, although only a rough correlation between estimated and experimental RRT is achieve, this proves to be sufficient in most cases to substantially reduce the length of the list of candidates.

9. "METABOLOMIC" AND CHEMOMETRIC DATA PROCESSING

The performance of the combination of the "metabolomic" and chemometric approaches in reducing the list of candidates was also assessed. The analysis was carried out against the 43 compounds already used in the evaluation of the chemometric approach (see chapter 7).

Example 1 regards the identification of the **bisoprolol**, for which a list of 2 candidates was retrieved. In this case, 1 peak for the dehydrogenation was found in the TIC, but it was not enough and the "metabolomic" method did not provide any reduction of the list. However, as previously illustrated, the application of the chemometric approach allowed to keep only one of the two candidates (bisoprolol).

Example 2 concerns the detection of **tenofovir**. The search against the PTRC database provided a list of 5 possible candidates. Cause no metabolites were found in the sample, the "metabolomic" approach could not be applied. Instead, the combination with the chemometric method allowed the elimination of 2 compounds from the list of candidates.

Example 3 refers to the determination of a citalopram metabolite: **didesmethylcitalopram**. From the PTRC database, a list of 6 candidates was achieved. The search in the TIC of the predefined set of biotransformations provided the identification of 1 peak corresponding to methylation. The acquired data were matched with the analysis of the molecular structure for each candidate obtained by E-Dragon. In this case the "metabolomic" approach enabled the exclusion of 4 candidates. The remaining compounds were: lophine and didesmethylcitalopram. By

applying the chemometric approach lophine could be excluded and didesmethylcitalopram was the only compound compatible after the combined application of the "metabolomic" and chemometric approaches.

Example 4 describes the case of **methadone**. The initial list of candidates included 10 isobaric compounds. The search of metabolites in the sample did not provide any results and therefore the "metabolomic" approach did not achieve any reduction in the list. The chemometric approach enabled the exclusion of 1 compound only.

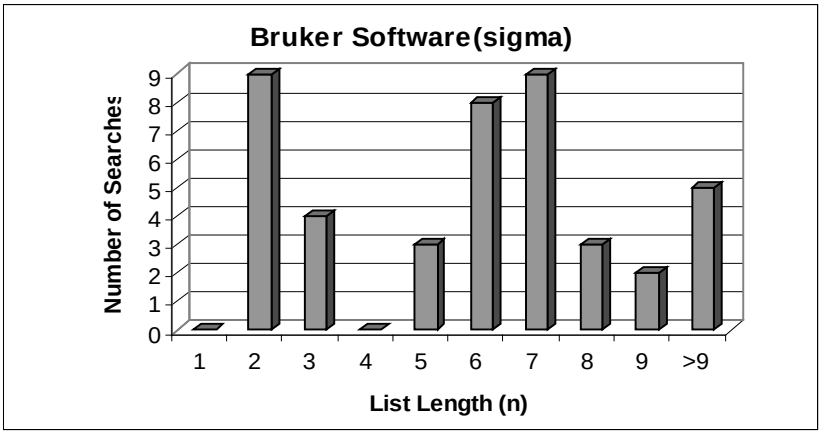
The application of the combined approaches on the 43 compounds allowed an overall reduction of the mean list length (MLL) from 6.02 ± 3.49 (median 6, range 2-21) to 2.14 ± 1.63 (median 2, range 1-9), that is more than what was obtained by either of the two approaches alone (Table 13)

A graphical evaluation of the performances of the metabolomic and chemometric approaches as well as of their combination is illustrated in Figure 13, where A shows the distribution of the list lengths for the 43 compounds as obtained by the Bruker software after searching for compatible exact mass/Isotopic patterns, and B, C, and D show the same distribution after applying the metabolomic, chemometric, and combined approaches, respectively (Figure 13).

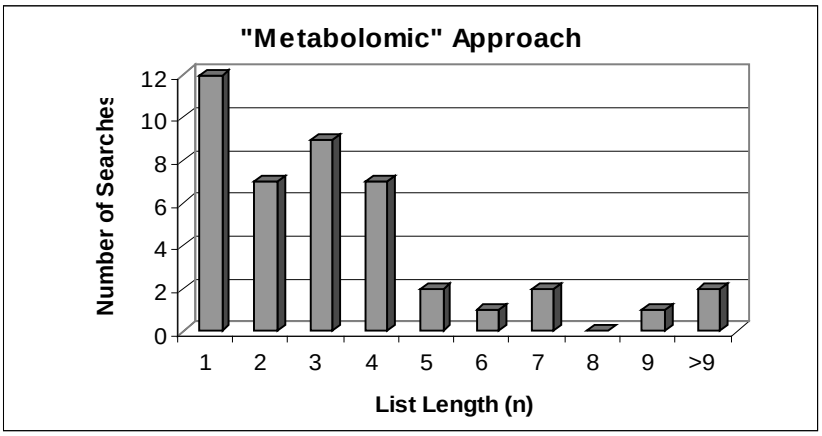
Table 13. *The mean, standard deviation, median, minimum and maximum calculated for each approach are also reported.*

List Length	Bruker (sigma)	Metabolomic Approach	Chemometric Approach	Metabolomic and Chemometric Approach
Mean	6.02	3.42	3.09	2.14
Standard Deviation	3.49	3.03	2.03	1.63
Median	6	3	2	2
MIN	2	1	1	1
MAX	21	17	9	9
N° Searches	43	43	43	43
Correct Candidate eliminated	0	0	0	0
List Length (No. Candidtes)				
1	0	12	7	19
2	9	7	17	12
3	4	9	6	8
4	0	7	5	1
5	3	2	1	0
6	8	1	3	2
7	9	2	2	0
8	3	0	1	0
9	2	1	1	1
>9	5	2	0	0
N° Searches (Total)	43	43	43	43

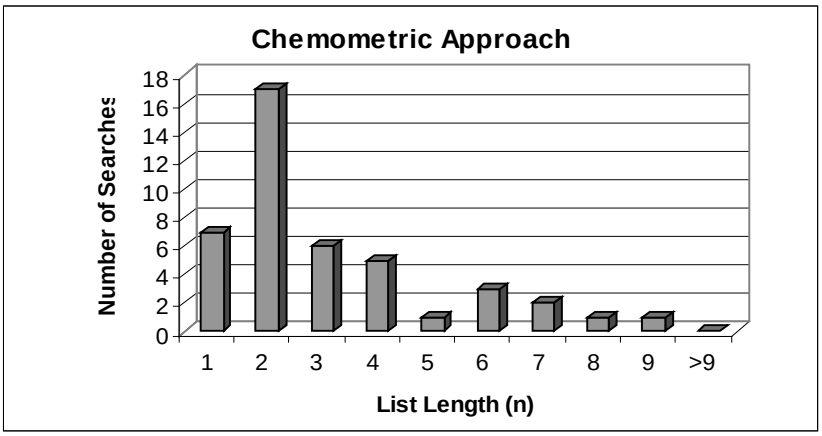
(A)



(B)



(C)



(D)

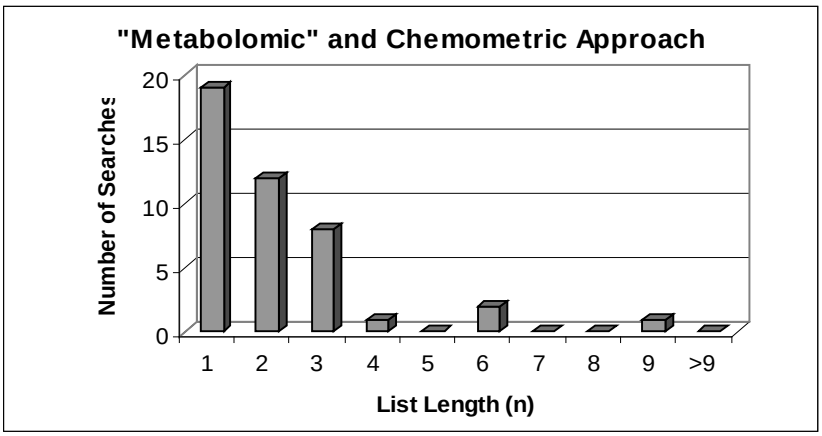


Figure 13. *Graphical description of the distribution of the compounds number in the candidates list related to the number of search for each evaluated approach.*

9.1. DISCUSSION

The simultaneous application of the "metabolomic" and the chemometric approach has highlighted the complementary features of these two strategies. Examples 1 and 2 (in the section 8) describing the search for bisoprolol and tenofovir respectively, have shown that even if the metabolomic approach did not provide any reduction of the candidates list, a further support could be achieved with the chemometric approach. In fact, in these two cases, the substantial lowering of the list was obtained by the evaluation of the molecular properties of each compound. Example 3 is the most striking case of the fundamental contribution to the search of an unknown compound that the two procedures could give when together applied. In this example the MLL was reduced from 6 to only 2 candidates by the "metabolomic" method and further shortened to only 1 candidate by the chemometric method allowing the unequivocal determination of the target compound (didesmethylcitalopram). On the contrary, even the combined application of the two strategies was not able to significantly reduce the list of candidates in the case of methadone.

However, data shown in Table 13 demonstrate the effectiveness of the combination of the two strategies in~~to~~ significantly shortening the list of candidates. In fact, both the "metabolomic" and the chemometric approach allow to reduce the MLL almost of the half, but their combination further improves the final result (from a MLL of 6.02 ± 3.49 to a MLL of 2.14 ± 1.63 after application of both the procedures). This point is extremely evident in the (Figure 13), where the bars of the histograms pertaining to the "metabolomic" (Figure 13 B), chemometric approach (Figure 13 C) as well as to the combination of the two methods (Figure 13 D) are shifted to the left as compared

to the data before the two data processing strategies (Figure 13 A), indicating the marked reduction in the MLL.

10. CONCLUSIONS

The availability of a wide PTRC database of compound names, chemical formula and exact masses, represents a strong tool for purposes of toxicological screening, broadening the number of identifiable compounds from a few 100s (by the library search approach or the use of in-house chemical databases) (Sauvage F.L. et al., 2006) to more than 50.000, thus reducing the number of false negatives. However, by widening the number of reference compounds included in the database, a higher number of false positives (compounds isobaric to the target), is obtained. Owing to this reason, additional tools for the reduction of the length of the list of candidates are essential. In this context the "metabolomic" approach and the chemometric approach here described proved to represent valid strategies for the achievement of this goal.

Studies on metabolomics have been introduced not only in the field of drug discovery and development (Baranczewski P. et al., 2006), but recently also for general unknown screening purposes (Pelander A. et al., 2009), although directed to the prediction and identification of metabolites of a known substance with the aim to add their exact masses and retention times in an in-house database including PTRC and their relative metabolites. This database currently consists of 830 compounds. This study describes for the first time the use and efficacy of a combined metabolomic and chemometric approach in the identification of unknown compounds using a large reference database (over 50.000).

Although the overall search procedure is well defined and each step of it has been automated, the future aim of our study will be the full automation of data processing, which implies to establish a communication between different softwares (*i.e.* Bruker

Data Analysis, E-Dragon Software, PubChem Compund database, and Windows Excel). Our future efforts will be directed to achieve this task.

REFERENCES

Aebi B., Bernhard W. Advances in the use of mass spectral libraries for forensic toxicology. *J Anal Toxicol* 2002; 26: 149-156.

Applications of LC-MS in toxicology. Polettini A., Ed.; Pharmaceutical Press: London, U.K., 2006.

Baczek T., Kaliszan R., Novotna K., Jandera P. Comparative characteristics of HPLC columns based on quantitative structure-retention relationships (QSRR) and hydrophobic-subtraction model. *J Chromatogr A* 2005; 1075: 109-115.

Baranczewski P., Stankzak A., Kautiainen A., Sandin P., Edlund P.O. Introduction to early in vitro identification of metabolites of new chemical entities in drug discovery and development. *Pharmacol Rep* 2006; 58:341-352.

Bogusz M.J. Hyphenated liquid chromatographic techniques in forensic toxicology. *J Chromatogr B* 1999; 733: 65-91.

Bogusz M.J., Maier R.D., Kruger K.D., Webb K.S., Romeril J., Miller M.L. Poor reproducibility of in-source collisional atmospheric pressure ionization mass spectra of toxicologically relevant drugs. *J Chromatogr A* 1999; 844: 409-418.

Boone C.M., Ensing K. Is capillary electrophoresis a method of choice for systematic toxicological analysis?. *Clin Chem Lab Med* 2003; 41: 773-781.

Boone C.M., Franke J.P., De Zeeuw R.A., Ensing K. Evaluation of capillary electrophoretic techniques towards systematic toxicological analysis. *J Chromatogr A* 1999; 838: 259-272.

D'Archivio A.A., Ruggieri F., Mazzeo P., Tettamanti E. Modelling of retention of pesticides in reverse-phase high-performance liquid chromatography: quantitative

structure-retention relationships based on solute quantum-chemical descriptors and experimental (solvatochromic and spin-probe) mobile phase descriptors. *Anal Chim Acta* 2007; 593: 140-151.

Deineka V.I. Relative retention analysis in reversed-phase HPLC: the determination, use, and role of the parameter Log*P* in the n-octanol-water system. *Russ J Phys Chem* 2006; 80: 950-954.

De Zeeuw R.A. Recent developments in analytical toxicology: for better or for worse. *Toxicol Lett* 1998; 102-103: 103-108.

Dresen S., Ferreiros N., Gnann H., Zimmermann R., Weinmann W. Detection and identification of 700 drugs by multi-target screening with a 3200 Q TRAP LC-MS/MS system and library searching. *Anal Bioanal Chem* 2010; article in press.

Dresen S., Kempf J., Weinmann W. Electrospray-ionization MS/MS library of drugs as database for method development and drug identification. *Forensic Sci Int* 2006; 161: 86-91.

Drummer O.H. Chromatographic screening techniques in systematic toxicological analysis. *J Chromatogr B* 1999; 733: 27-45.

Drummer O.H. Requirements for bioanalytical procedures in postmortem toxicology. *Anal Bioanal Chem* 2007; 388: 1495-1503.

Elliott S.P., Hale K.A. Applications of an HPLC-DAD drug-screening system based on retention indices and UV spectra. *J Anal Toxicol* 1998; 22: 279-289.

Etxebarria N., Zuloaga O., Olivares M., Bartolome L.J., Navarro P. Retention-time locked methods in gas chromatography. *J Chromatogr A* 2008; 1216: 1624-1629.

Fang L., Demee M., Cournoyer J., Sierra T., Young C., Yan B. Parallel high-throughput accurate mass measurement using a nine-channel multiplexed

electrospray liquid chromatography ultraviolet-time-of-flight mass spectrometry system. *Rapid Commun Mass Spectrom* 2003; 17: 1425-1432.

Ferrer I., Garcia-Reyes J.F., Mezcua M., Thurman E.M., Fernandez-Alba A.R. Multi-residue pesticide analysis in fruits and vegetables by liquid chromatography-time-of-flight mass spectrometry. *J Chromatogr A* 2005; 1082: 81-90.

Ferrer I., Thurman E.M. Multi-residue method for the analysis of 101 pesticides and their degradates in food and water samples by liquid chromatography/time-of-flight mass spectrometry. *J Chromatogr A* 2007; 1175: 24-37.

Ferrer I., Thurman E.M., Fernandez-Alba A.R. Quantitation and accurate mass analysis of pesticides in vegetables by LC/TOF-MS. *Anal Chem* 2009; 77: 2818-2825.

Gaillard Y., Pepin G. Use of high-performance liquid chromatography with photodiode-array UV detection for the creation of a 600-compounds library application to forensic toxicology. *J Chromatogr A* 1997; 763: 149-163.

Gangl E., Utkin I., Gerber N., Vouros P. Structural elucidation of metabolites of ritonavir and indinavir by liquid chromatography-mass spectrometry. *J Chromatogr A* 2002; 974: 91-101.

Gergov M., Boucher B., Ojanpera I., Vuori E. Toxicological screening of urine for drugs by liquid chromatography/time-of-flight mass spectrometry with automated target library search based on elemental formulas. *Rapid Commun Mass Spectrom* 2001; 15: 521-526.

Gergov M., Ojanpera I., Vuori E. Simultaneous screening for 238 drugs in blood by liquid chromatography-ionspray tandem mass spectrometry with multiple-reaction monitoring. *J Chromatogr B* 2003; 795: 41-53.

Gergov M., Weinmann W., Meriluoto J.,Uusitalo J.,Ojanpera I. Comparison of product ion spectra obtained by liquid chromatography/triple-quadrupole mass spectrometry for library search. *Rapid Commun Mass Spectrom* 2004; 18: 1039-1046.

Gottardo R., Fanigliulo A., Bortolotti F., De Paoli G., Pascali J.P., Tagliaro F. Broad-spectrum toxicological analysis of the hair based on capillary electrophoresis-time-of-flight mass spectrometry. *J Chromatogr A* 2007; 1159: 190-197.

Gottardo R., Polettini A., Sorio D., Pascali J.P., Bortolotti F., Liotta E., Tagliaro F. Capillary zone electrophoresis (CZE) coupled to time-of-flight mass spectrometry (TOF-MS) applied to the analysis of illicit and controlled drugs in blood. *Electrophoresis* 2008; 29: 4078-4087.

Gratz S.R., Gamble B.M., Flurer R.A. Accurate mass measurement using Fourier transform ion cyclotron resonance mass spectrometry for structure elucidation of designer drug analogs of tadalafil, vardenafil and sildenafil in herbal and pharmaceutical matrices. *Rapid Commun Mass Spectrom* 2006; 20: 2317-2327.

Hayashida M., Takino M., Terada M., Kurisaki E., Kudo K.,Ohno Y. Time-of-flight mass spectrometry (TOF-MS) exact mass database for benzodiazepine screening. *Leg Med* 2009; 11: S423-S425.

Heiger D. Principles of capillary electrophoresis. In high performance capillary electrophoresis; Agilent Technologies; Germany, 2000, pp. 16-44.

Hendickson C.L., Emmett M.R. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Annu Rev Phys Chem* 1999; 50: 517-536.

Herrero M., Garcia-Canas V., Simo C., Cifuentes A. Recent advances in the application of capillary electromigration methods for food analysis and foodomics. *Electrophoresis* 2010; 31: 205-228.

Hill D.W., Kertesz T.M., Fontaine D., Friedman R., Grant D.F. Mass spectral metabonomics beyond elemental formula: chemical database querying by matching

experimental with computational fragmentation spectra. *Anal Chem* 2009; 80(14): 5574-5582.

Hill D.W., Kind A.J. Reversed-phase solvent-gradient HPLC retention indexes of drugs. *J Anal Toxicol* 1994; 18: 233-242.

Hogendoorn E.A., Hoogerbrugge R., Baumann R.A., Meiring H.D., De Jong A.P., Van Zoonen P. Screening and analysis of polar pesticides in environmental monitoring programmes by coupled-column liquid chromatography and gas chromatography-mass spectrometry. *J Chromatogr A* 1996; 754: 49-60.

Holcapek M., Kolarova L., Nobilis M. High-performance liquid chromatography-tandem mass spectrometry in the identification and determination of phase I and phase II metabolites. *Anal Bioanal Chem* 2008; 391: 59-78.

Hopfgartner G., Husser C., Zell M. High-throughput quantification of drugs and their metabolites in biosamples by LC-MS/MS and CE-MS/MS: possibilities and limitations. *Ther Drug Monit* 2002; 24: 134-143.

Ibanez M., Sancho J.V., Pozo O.J., Niessen W., Hernandez F. Use of quadrupole time-of-flight mass spectrometry in the elucidation of unknown compounds present in environmental water. *Rapid Commun Mass Spectrom* 2005; 19: 169-178.

Jansen R., Lachatre G., Marquet P. LC-MS/MS systematic toxicological analysis: comparison of MS/MS spectra obtained with different instruments and settings. *Clin Biochem* 2005; 38: 362-372.

Kind T., Fiehn O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* 2006; 7: 234-243.

Klampfl C.W. CE with MS detection: a rapidly developing hyphenated technique. *Electrophoresis* 2009; 30: S83-S91.

Koh H.L., Yau W.P., Ong P.S.,Hegde A. Current trends in modern pharmaceutical analysis for drug discovery. *Drug Discov Today* 2003; 8: 889-897.

Kolmonen M., Leinonen A., Pelander A., Ojanpera I. A general screening method for doping agents in human urine by solid phase extraction and liquid chromatography/time-of-flight mass spectrometry. *Anal Chim Acta* 2007; 585: 94-102.

Kronstrand R., Josefsson M. Quantification using LC-MS. In Applications of LC-MS in Toxicology; Polettini A., Ed.; Pharmaceutical Press: London, U.K., 2006, pp. 43-70.

Lacorte S., Fernandez-Alba A.R. Time of flight mass spectrometry applied to the liquid chromatographic analysis of pesticides in water and food. *Mass Spectrom Rev* 2006; 25: 866-880.

Lafaye A., Junot C., Gall B.R., Fritsch P., Tabet J.C., Ezan E. Metabolite profiling in rat urine by liquid chromatography/electrospray ion trap mass spectrometry. Application t the study of heavy metal toxicity. *Rapid Commun Mass Spectrom* 2003; 17: 2541-2549.

Lambert W.E., Meyer E., Leenheer A.P.D. Systematic toxicological analysis of basic drugs by gradient elution of an alumina-based HPLC packing material under alkaline conditions. *J Anal Toxicol* 1995; 19: 73-78.

Lambert W.E. Pitfalls in LC-MS(-MS) analysis. Paper presented at the Joint GTFCh/TIAFT Symposium “Analytical pitfalls and trands in clinical and forensic toxicology” during the Analytica Conference, May 12th, 2004 in Munich.

Law B., Temesi D. Factors to consider in development of generic bioanalytical high-performance liquid chromatographic-mass spectrometric methods to support drug discovery. *J Chromatogr B* 2000; 748: 21-30.

Lei B., Li S., Xi L., Li J., Liu H., Yao X. Novel approaches for retention time prediction of oligonucleotides in ion-pair reversed-phase high-performance liquid chromatography. *J Chromatogr A* 2009; 1216: 4434-4439.

Loukas Y.L. Artificial neural networks in liquid chromatography: efficient and improved quantitative structure-retention relationship models. *J Chromatogr A* 2000; 904: 119-129.

Manetto G., Tagliaro F., Crivellente F., Pascali V.L., Marigo M. Field-amplified sample stacking capillary zone electrophoresis applied to the analysis of opiate drugs in hair. *Electrophoresis* 2000; 21: 2891-2898.

Marquet P. Is LC-MS suitable for a comprehensive screening of drugs and poisons in clinical toxicology? *Ther Drug Monit* 2002; 24: 125-133.

Marquet P. Systematic toxicological analysis with LC-MS. In Applications of LC-MS in Toxicology; Polettini A., Ed.; Pharmaceutical Press: London, U.K., 2006, pp. 111-130.

Martnez-Bueno M.J., Agera A., Gmez M.J., Hernando M.D., Garcia-Reyes J.F., Fernandez-Alba A.R. Application of liquid chromatography/quadrupole-linear ion trap mass spectrometry and time-of-flight mass spectrometry to the determination of pharmaceuticals and related contaminants in wastewater. *Anal Chem* 2007; 79: 9372-9384.

Mass spectral and GC data of drugs, poisons, pesticides, pollutants and their metabolites, 3rd ed.; Maurer H.H., Pfleger K., Weber A., Eds.; Wiley-VCH: Weinheim, Germany, 2007.

Maurer H.H., Bickeboeller-Friedrich J. Screening procedure for detection of antidepressants of the selective serotonin reuptake inhibitor type and their metabolites in urine as part of a modified systematic toxicological analysis procedure using gas chromatography-mass spectrometry. J *Anal Toxicol* 2000; 24: 340-347.

Maurer H.H. Current role of liquid chromatography-mass spectrometry in clinical and forensic toxicology. *Anal Bioanal Chem* 2007; 388: 1315-1325.

Maurer H.H. Hyphenated mass spectrometric techniques – indispensable tools in clinical and forensic toxicology and in doping control. *J Mass Spectrom* 2006; 41: 1399-1413.

Maurer H.H. Position of chromatographic techniques in screening for detection of drugs or poisons in clinical and forensic toxicology and/or doping control. *Clin Chem Lab Med* 2004; 42: 1310-1324.

Mortier K.A., Zhang G.F., Van Peteghem C.H., Lambert W.E. Adduct formation in quantitative bioanalysis: effect of ionization condidtions on paclitaxel. *J Am Soc Mass Spectrom* 2004; 15: 585-592.

Mortishire-Smith R.J., Castro-Perez J.M., Yu K., Shockcor J.P., Goshawk J., Hartshorn M.J., Hill A. Generic dealkylation: a tool for increasing the hit-rate of metabolite rationalization, and automatic customization of mass defect filters. *Rapid Commun Mass Spectrom* 2009; 23: 939-948.

Mottram D.R. Banned drugs in sport: does the International Olympic Committee (IOC) list need updating? *Sports Med* 1999; 27: 1-10.

Mueller C.A., Weinmann W., Dresen S., Shreiber A., Gergov M. Development of a multi-target screening analysis for 301 drugs using a QTrap liquid chromatography/tandem mass spectrometry system and automated library searching. *Rapid Commun Mass Spectrom* 2005; 19:1332-1338.

National Institute of Standards and Technology. NIST Chemistry Webbook. http://webbook.nist.gov/chemistry/ (accessed December 2007).

National Instiutute of Standard and Technology. NIST/EPA/NIH Mass Spectral Library. http//www.nist.gov/data/nist1a.htm (accessed December 2007).

Nielen M.W.F., van Engelen M.C., Zuiderent R., Ramaker R. Screening and confirmation criteria for hormone residue analysis using liquid chromatography accurate time-of-flight, Fourier transform ion, cyclotron resonance and orbitrap mass spectrometry techniques. *Anal Chim Acta* 2007; 586: 122-129.

Ojanpera I., Pelander A., Laks S., Gergov M., Vuori E., Witt M. Application of accurate mass measurement to urine drug screening. *J Anal Toxicol* 2005; 29: 34-40.

Ojanpera S., Pelander A., Pelzing M., Krebs I., Vuori E., Ojanpera I. Isotopic pattern and accurate mass determination in urine drug screening by liquid chromatography/time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2006; 20: 1161-1167.

Patel B.N., Sharma N., Sanyal M., Shrivastav P.S. High throughput and sensitive determination of trazodone and its primary metabolite, m-chlorphenylpiperazine, in human plasma by liquid chromatography-tandem mass spectrometry. *J Chromatogr B* 2008; 871: 44-54.

Pelander A., Tyrkko E., Ojanpera I. In silico methods for predicting metabolism and mass fragmentation applied to quetiapine in liquid chromatography/time-of-flight mass spectrometry urine drug screening. *Rapid Commun Mass Spectrom* 2009; 23: 506-514.

Pelzing M., Neusub C., Macht M. Accurate mass and TIP^{TM}: two dimensions for elemental composition determination. LC-GC Eur 2004; 17: 38-39.

Peters R.J.B., Oosterink J.E., Stolker A.A.M., Georgakopoulos C., Nielen M.W.F. Generic sample preparation combined with high resolution liquid chromatography-time-of-flight mass spectrometry for unification of urine screening in doping-control laboratories. *Anal Bioanal Chem* 2010; accepted for publication.

Plumb R.S., Stumpf C.L., Granger J.H., Castro-Perez J., Naselden J.N., Dear G.J. Use of liquid chromatography/time-of-flight mass spectrometry and multivariate

statistical analysis shows promise for the detection of drug metabolites in biological fluids. *Rapid Commun Mass Spectrom* 2003; 17: 2632-2638.

Polettini A., Gottardo R., Pascali J.P., Tagliaro F. Implementation and performance evaluation of a database of chemical formulas for the screening of pharmaco/toxicologically relevant compounds in biological samples using electrospray ionization-time-of-flight mass spectrometry. *Anal Chem* 2008; 80: 3050-3057.

Polettini A., Groppi A., Vignali A., Montagna M. Fully-automated systematic toxicological analysis of drugs, poisons, and metabolites in whole blood, urine, and plasma by gas-chromatography-full scan mass spectrometry. *J Chromatogr B* 1998; 713: 265-279.

Polettini A. Systematic toxicological analysis of drugs and poisons in biosamples by hyphenated chromatographic and spectroscopic techniques. *J Chrom B* 1999; 733: 47-63.

Politi L., Groppi A., Polettini A. Ionization, ion separation and ion detection in LC-MS. In Applications of LC-MS in Toxicology; Polettini A., Ed.; Pharmaceutical Press: London, U.K., 2006, pp. 1-22.

Pommier F., Frigola R. A simple, rapid and sensitive method for simultaneous determination of rivastigmine and its major metabolite NAP 226-90 in rat brain and plasma by reversed-phase liquid chromatography coupled to electrospray ionization mass spectrometry. *J Chromatogr B: Anal Technol Biomed Life Sci* 2003; 18: 160-166.

Put R., Heyden Y.V. Review on modeling aspects in reversed-phase liquid chromatographic quantitative structure-retention relationship. *Anal Chim Acta* 2007; 602: 164-172.

Rekker R., Mannhold R., ter Laak A.M. On the reliability of calculated Log*P*-values: Rekker, Hansch/Leo and Suzuki approach. *Quant Struct –Act Relat* 1993; 12: 152-157.

Remko M., Swart M., Bickelhaupt M. Theoretical study of structure, pKa, lipophilicity, solubility, absorption, and polar surface area of some centrally acting antihypertensive. *Bioorg Med Chem* 2006; 14: 1715-1728.

Rivier L. Criteria for the identification of compounds by liquid chromatography-mass spectrometry and liquid chromatography-multiple mass spectrometry in forensic toxicology and doping control. *Anal Chim Acta* 2003; 492: 69-82.

Rivier L. Identification and confirmation criteria for LC-MS. In Applications of LC-MS in Toxicology; Polettini A., Ed.; Pharmaceutical Press: London, U.K., 2006, pp. 97-109.

Rourick R.A., Volk K.J., Klohr S.E., Spears T., Kerns E.H., Lee M.S. Predictive strategy for the rapid structure elucidation of drug degradants. *J Pharm Biomed Anal* 1996; 14: 1743-1752.

Sancho J.V., Pozo O.J., Ibanez M., Hernandez F. Potential of liquid chromatography/time-of-flight mass spectrometry for the determination of pesticides and transformationproducts in water. *Anal Bioanal Chem* 2006; 386: 987-997.

Sauvage F.L., Saint-Marcoux F., Duretz B., Deporte D., Lachatre G., Marquet P. Screening of drugs and toxic compounds with liquid chromatography-linear ion trap tandem mass spectrometry. *Clin Chem* 2006; 52: 1735-1742.

Schauer N., Steinhauser D., Strelkov S., Schomburg D., Allison G., Moritz T., Lundgren K., Roessner-Tunali U., et al. GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett* 2005; 579: 1332-1337.

Schrader W., Klein H.W. Liquid chromatography/Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR MS): an early overview. *Anal Bioanal Chem* 2004; 379: 1013-1024.

Shamsi S.A., Miller B.E. Capillary electrophoresis-mass spectrometry: recent advances to the analysis of small achiral and chiral solutes. *Electrophoresis* 2004; 25: 3927-3961.

Slater B., McCormack A., Avdeef A., Comer J.E.A. pH-Metric Log*P* 4. Comparison of partition coefficients determined by shake-flask, HPLC, and potentiometric methods. *J Pharmaceut Sci* 1994; 83: 1980-1983.

Soriano T., Jurado C., Menendez M., Repetto M. Improved solid-phase extraction method for systematic toxicological analysis in biological fluids. *J Anal Toxicol* 2001; 25: 137-143.

Staub A., Schappler J., Rudaz S., Veuthey J.L. CE-TOF/MS: fundamental concepts, instrumental considerations and applications. *Electrophoresis* 2009; 30: 1610-1623.

Stimpfl T., Vycudilik W. Identification of the general unknown. Application of mass selective detectors in forensic toxicology. *J Anal Toxicol* 2000; 24: 32-36.

Strano-Rossi S., Malaioni F., Rossi F., Botrè F. Rapid screening of drugs of abuse and their metabolites by gas chromatography/mass spectrometry: application to urinalysis. *Rapid Commun Mass Spectrom* 2005; 19: 1529-1535.

Tagliaro F., Antonioli C., De Battisti Z., Ghielmi S., Marigo M. Reversed-phase high-performance liquid chromatographic determination of cocaine in plasma and human hair with direct fluorimetric detection. *J Chromatogr A* 1994; 674: 207-215.

Tetko I.V. Computing chemistry on the web. *Drug Discov Today* 2005; 10: 1497-1500.

Tetko I.V., Gasteiger J., Todeschini R., Mauri A., Livingstone D., Ertl P., Palyulin V.A., Eugene V., et al. Virtual computational chemistry laboratory – design and description. *J Comput Aid Mol Des* 2005; 19: 453-463.

Tetko I.V., Tanchuk V.Y., Villa A.E.P. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J Chem Inf Comput Sci* 2001; 41:1407-1421.

The Merck index. O'Neil M., Heckelman P.E., Koch C.B., Roman K.J., Kenny C.M., Eds.; Merck & Company Incorporated: Whitehouse Station, N.J., 2006.

Thurman E.M., Ferrer I., Fernandez-Alba A.R. Matching unknown empirical formulas to chemical structure using LC/MS TOF accurate mass and database searching: example of unknown pesticides on tomato skins. *J Chromatogr A* 2004; 1067: 127-134.

Thurman E.M., Ferrer I., Malato O., Fernandez-Alba A.R. Feasibility of LC/TOFMS and elemental database searching as a spectral library for pesticides in food. *Food Addit Contam* 2006; 23: 1169-1178.

Thurman E.M., Ferrer I., Pozo O.J., Sancho J.V., Hernandez F. The even-electron rule in electrospray mass spectra of pesticides. *Rapid Commun Mass Spectrom* 2007; 21: 3855-3868.

United States National Library of Medicine, National Institutes of Health, National Center for Biotechnology Information. Medical Subjects Headings Database. http//www.nlm.nih.gov/mesh/mBrowser.html (accessed December 2007).

United States National Library of Medicine, National Institutes of Health, National Center for Biotechnology Information. PubChem Compound. http:// pubchem.ncbi.nlm.nih.gov/ (accessed December 2007).

United States National Library of Medicine, National Institutes of Health, Specialized Information Services ChemIDplus Advanced http//chemsis.nlm.nih.gov/chemidplus/ (accessed December 2007).

VCCLAB. Virtual Computational Chemistry Laboratory. http//www.vcclab.org, 2005.

Vu H., Pham N.B., Quinn R.J. Direct screening of natural product extracts using mass spectrometry. *J Biomol Screen* 2008; 13: 265-275.

Wang R., Fu Y., Lai L. A new atom-additive method for calculating partition coefficients. *J Chem Inf Comput Sci* 1997; 37: 615-621.

Watzig H., Gunter S. Capillary electrophoresis-a high performance analytical separation technique. *Clin Chem Lab Med* 2003; 41: 724-738.

Weinmann W., Lehmann N., Muller C., Wiedemann A., Svoboda M. Identification of lorazepam and sildenafil as examples for the application of LC/ionspray-MS and MS-MS with mass spectra library searching in forensic toxicology. *Forensic Sci Int* 2000; 113: 339-344.

Weinmann W., Stoertzel M., Vogt S., Wendt J. Tune compounds for electrospray ionisation/in-source collision-induced dissociation with mass spectral library searching. *J Chromatogr A* 2001; 926: 199-209.

Weinmann W., Wiedemann A., Eppinger B., Renz M., Svoboda M. Screening for drugs in serum by electrospray ionization/collision-induced dissociation and library searching. *J Am Soc Mass Spectrom* 1999; 10: 1028-1037.

Wiley. Wiley Registry of Mass Spectral Data. http//eu.wiley.com/Wiley-CDA/WileyTitle/productCd-0470047852, miniSiteCd-STMDB2.html (accessed December 2007).

Williams T.A., Riddle M., Morgan S.L., Brewer W.E. Rapid gas chromatographic analysis of drugs of forensic interest. *J Chromatogr Sci* 1999; 37: 210-214.

Zhang H., Zhang D., Ray K., Zhu M. Mass defect filter technique and its applications to drug metabolite identification by high resolution mass spectrometry. *J Mass Spectrom* 2009; 44: 999-1016.

Zhang H., Zhu M., Ray K., Ma L., Zhang D. Mass defect profiles of biological matrices and general applicability of mass defect filtering for metabolite detection. *Rapid Commun Mass Spectrom* 2008; 22: 2082-2088.

Zhu M., Ma L., Zhang D., Ray K., Zhao W., Humphreys G.W., Skiles G., Sanders M., Zhang H. Detection and characterization of metabolites in biological matrices using mass defect filtering of liquid chromatography/high resolution mass spectrometry data. *Drug Metab Dispos* 2006; 34:1722-1733.

Acknowledgements

Special thanks to:

Prof. Aldo Polettini and Prof. Franco Tagliaro,

Dr. Federica Bortolotti, Dr. Rossella Gottardo, Dr. Jennifer Pascali,

Dr. Daniela Sorio, Dr. Anna Bertaso,

Mrs. Giovanna Carli, Mrs. Mara Mazzola