USE OF SOFT AND HARD MS IONIZATION TECHNIQUES FOR UNKNOWN COMPOUNDS ELUCIDATION BY GC-TOF MS

Tania Portolés[1], Elena Pitarch[1], Francisco J. López[1], Félix Hernández[1*], W.M.A. Niessen[2].

[1.] Research Institute for Pesticides and Water, University Jaume I, 12071 Castellón, Spain.

[2.] hyphen MassSpec, de Wetstraat 8, 2332 XT Leiden, the Netherlands

Tel: 34-964-387366, Fax: 34-964-387368, E-mail: felix.hernandez@qfa.uji.es

ABSTRACT

Investigation of trace level non-target compounds by GC-MS often is a challenging task that requires powerful software tools to detect the unknown components, to obtain the deconvoluted mass spectra, and to interpret the data if no acceptable library match is obtained. In this paper, the complementary use of EI and CI is investigated in combination with GC-TOF MS for the elucidation of organic non-target (micro)contaminants in water samples. Based on accurate mass measurement of the molecular and fragment ions from the TOF MS, empirical formulae were calculated. Isotopic patterns, carbon number prediction filter and nitrogen rule were used to reduce the number of possible formulae. The candidate formulae were searched in databases to find possible chemical structures. Selection from possible structure candidates was achieved using information on substructures and observed neutral losses derived from the fragment ions. Four typical examples (bifenazate, boscalid, epoxiconazole, and fenhexamid) are used to illustrate the methodology applied and the various difficulties encountered in this process. Our results indicate that elucidation of unknowns cannot be achieved by following a standardized procedure, as both expertise and creativity are necessary in the process.

INTRODUCTION

Gas chromatography coupled to mass spectrometry (GC-MS) is one of the most powerful techniques for detection, identification and quantification of volatile and semi-volatile contaminants and residues in environmental, biological and food matrices. In these applications, electron ionization (EI) is the most widely used ionization technique. The ability of an EI source to produce highly reproducible fragmentation spectra allows obtaining valuable structural information on the molecules and the generation of large spectral libraries highly useful for qualitative analysis. This allows compound identification based on matching experimental spectra to mass spectral databases libraries. In addition, retention index matching is important to distinguish isomers. The identification process gains power when the *m/z* values of the ions in the EI spectrum are measured with high mass accuracy, as occurs when using high resolution time-of-flight mass spectrometry (TOF MS). Under these circumstances, the compounds identified by library matching can be confirmed by accurate mass measurements of the fragment ions and the molecular ion (if present in the EI spectrum) and can solve ambiguous results in library search [1]. The versatility of large libraries lays in the fact that EI mass spectra are comparable over a wide range of different types of mass spectrometers from different vendors, although quadrupoles may be tune to preferentially transmit high *m/z* ions, which thus result in slightly different ion abundance from those in TOF MS spectra [2]. This fact together with the high degree of fragmentation normally generated by this “hard”

ionization technique can be a trouble when the EI experimental spectrum does not yield a conclusive library match. This situation may occur for many compounds (new emerging contaminants, transformation products, not regulated compounds, etc) that are not included in available libraries, which makes the identification process more difficult. Under these circumstances, alternative ways of identification are needed.

The presence of the molecular ion in the mass spectrum, especially if measured at accurate mass, is a valuable tool as it provides indispensable information about the identity of the unknown compound. For these purpose, “soft” ionization techniques are required that produce spectra with less fragmentation and keep the molecule intact. Examples of “soft” ionization techniques are chemical ionization (CI), negative ion chemical-ionization (NICI), field ionization (FI), and atmospheric pressure chemical ionization (APCI) [2-5].

Once the molecular (adducts) ions have been identified, either from EI or CI data, accurate mass and isotope data can be used to calculate formulae. The exact mass differences between ions can also be used to search and/or confirm the identity of neutral lost and in some instances clarify fragmentation pathways. The difference in mass due to the loss of a specific functional group is often relatively small. Therefore, the formulae based on these mass differences are very specific and unambiguous because of the reduced number of combinations of elements possible. This allows unambiguous assignments of the losses within the spectrum. Conversely, this information can in most cases be used to confidently determine the formula of the

molecular (adduct) ion of an unknown analyte.

Not many authors have reported examples on elucidation of compounds when their library mass spectra are not available. For finding candidate structures for unknowns, the group of Schymanski reported the use of database searches and structure generation, together with the incorporation of analytical and Effect Direct Analysis (EDA)-specific information in computer-based methods. This allowed reducing the number of candidate structures [6-9]. Within this process, accurate masses measurements facilitated the prediction of elemental compositions of a wider range of unknowns [6].

GC-TOF MS has been used for the identification of unknown compounds detected in extracts of well water. These first experiments showed that the GC-TOF MS instrument was not as powerful for determining ion compositions as double-focusing mass spectrometers, perhaps due to the fact that resolution of early TOF instruments (~5000) was by far not as good as that of double focussing-sector instruments (generally >10.000) [3]. More recently, a method based on the use of GC-TOF MS with an APCI source has been optimized for 31 compounds (amino acid, organic acids, alcohols, xanthines, etc) for which the standard mixture was available [4]. It was applied to human cerebrospinal fluid (CSF) samples for metabolic profiling. More than 300 compounds with different isotopic features were determined in the CSF samples. The identity of some of those peaks could be corroborated by the standards included in the mixture (comparing retention time, *m/z* value, and isotopic pattern of standard an samples). When no standard was available, only the *m/z* value and isotope pattern was used to derive the molecular formulae of the analytes present in the CSF. The combination of hard and soft ionization techniques for elucidation purposes has been described in a few papers. The use of FI in combination with EI was evaluated for targeted polycyclic hydrocarbons and several other model compounds [5]. Despite the lower ion yields obtained by FI, its ability to produce molecular ions and chromatograms with good S/N was impressive, especially in

combination with the ability to accurately measure the molecular mass using TOF MS [5].

In metabolomic applications with GC-(EI)MS, the low abundance of the molecular ions normally impedes the calculation of formulae for the identification of unknowns. On changing the beam-steering voltage of the ion source, the relative abundances of molecular ions at 70 eV were increased up to ten-fold for alkanes, fatty acid methyl esters and trimethylsilylated metabolites, concomitant with 2-fold absolute increases in ion intensities [2]. Next, the abundance, mass accuracy and isotope ratio accuracy of molecular species in EI has been compared with those in CI with methane as reagent gas under high-mass tuning. When constraining lists of calculated elemental compositions by chemical and heuristic rules using the Seven Golden Rules algorithm and PubChem queries, the correct formula was retrieved as top hit in 60% of the cases and within the top-3 hits in 80% of the cases [2].

The Seven Golden Rules, developed by Kind et al [2], enable an automatic exclusion of molecular formulas which are either wrong or which contain unlikely high or low number of elements. They are a set of heuristic rules for element composition calculations, including, among others, Senior and Lewis rules, element ratio rules and an isotopic abundance matching filter [10]. The advances in structure elucidation of small molecules using mass spectrometry have been recently reviewed by Kind et. al [11]. This interesting review covers different soft and hard ionization techniques and figures of merit for modern mass spectrometers. Also, mass spectral data handling strategies and mass spectral fragmentation pathways are discussed and the importance of mass spectral library search algorithms is outlined. The current state of the software development for the advancement of structure elucidation of small molecules is also reviewed.

Other applications reported in the non-target field deal with the identification of impurities generated in organic synthesis or in flavor research using the accurate mass measurements provided by TOF MS. This allowed the elucidation of compounds that could not be identified when applying GC-quadrupole systems [12,13].

High-resolution (Q)TOF MS instruments have also been used in combination with LC for the identification of chemical formulas of small molecules in the screening of pharmaco-toxicologically relevant compounds and drugs [14,15], and in the elucidation process of unknowns in environmental samples [16].

In this paper, EI and CI sources have been applied for the elucidation of the identity of organic contaminants in water samples. The model compounds investigated corresponded to pesticides, which have been chosen because their mass spectra are not registered in the commercial library available in our laboratory. In this way, the elucidation procedure was applied treating these compounds as fully unknowns. The $[M+H]^+$ in methane positive CI spectrum was usually abundant and often represented the base peak of the spectrum. The degree of fragmentation of $[M+H]^+$ ions was much lower than under 70 eV EI conditions, as the extent of exothermicity of the protonation in CI is lower, resulting in internal energy minor than in EI.

EXPERIMENTAL

Reagents

Reference standards of pesticides were purchased from Dr. Ehrenstorfer (Augsburg, Germany). From solid reference standards, stock solutions (around 500 μg/mL) were prepared by dissolving reference standards in acetone and stored in a freezer at –20°C. Working solutions were prepared by diluting stock solutions in hexane for extract fortification and injection in the chromatographic system. Acetone (residue analysis), ethyl acetate, dichloromethane and hexane (ultra-trace quality) were purchased from Scharlab (Barcelona, Spain). About 500 mg Bond Elut cartridges C_{18} (Varian, Harbor City, CA, USA) were used for solid-phase extraction.

Instrumentation

For the GC instrumentation, an Agilent 6890N GC system (Palo Alto, CA, USA) equipped with an Agilent 7683 autosampler was coupled to a GCT time-of-flight mass spectrometer (Waters Corporation, Manchester, UK), operating in EI and CI modes. The instrument was operated under MassLynx version 4.1 (Waters Corporation)

The GC separation was performed using a fused silica HP-5MS capillary column with 30 m x 0.25 mm i.d. and a film thickness of 0.25 μm (J&W Scientific, Folson, CA, USA). The oven temperature was programmed as follows: 90ºC (1 min); 5ºC/min to 300ºC (2 min). Injector temperature was set to 280ºC. Splitless injections of 1 μL sample were carried out. Splitless time was set to 1 min applying a constant gas flow of 1 mL/min. Helium was used as carrier gas at 1 mL/min.

The interface temperature was set to 250ºC and the source temperatures were set to 250ºC

and 100 ºC for EI and CI source, respectively. Electron energy was 70 eV for EI and 100 eV for CI sources. Methane was used as a CI reagent gas, with a source preassure of $2e^{-4}$ mbar. A solvent delay of 3 minutes was selected. TOF MS was operated at 1 spectrum/ s acquiring the mass range *m/z* 50-650 and using a multi-channel plate voltage of 2800 V. TOF MS resolution was about 8500 (FWHM) at *m/z* 614. Heptacosa, used for the daily mass calibration, was injected via syringe in the reference reservoir at 30ºC for this purpose. Additionally, heptacosa was used as a lock mass correction for EI experiments (monitoring the ion with *m/z* 218.9856); tris-(trifluoromethyl)-triazine for positive CI experiments (monitoring the ion with *m/z* 286.0027); and chloropentafluorobenzene for negative CI experiments (monitoring the ion with *m/z* 201.9609). Methane was used as reagent gas in the CI source.

General methodology

250 mL of groundwater were passed through a 500 mg C_{18} solid-phase extraction cartridge previously conditioned. After loading the sample, cartridges were washed with 3 mL water, air-dried using vacuum for at least 15 min, and then eluted with 5 mL ethyl acetate:dichloromethane (50:50). The extract was evaporated to dryness under a gentle nitrogen stream at 40°C and redissolved in 0.5 mL hexane. The final extract obtained was spiked with a mixture of selected pesticides at a concentration of 1 µg/ml (adding 10µl of 50 µg/ml standard) and it was injected into the GC-TOF MS. Three different injections were carried out, one for each ionization mode employed (EI, positive CI, negative CI). Then, TOF MS full-acquisition data were processed, treating the sample as unknown, i.e., using non-target processing method [17-19].

Data processing

In the first place, EI data were processed in a non-target way by applying the ChromaLynx Application Manager, a module of MassLynx software. This software automatically detects peaks with a response over user-defined parameters, displays their deconvoluted mass spectra, searches them against the commercial nominal mass NIST02 library, and produces a hit list with positive matches (library match >700 was used as criterion). An Elemental Composition Calculator is applied to derive the five most likely formulae of up to five most intense ions in the experimental TOF MS spectrum. These fragment formulae are tested against the molecular formulae of the top-five library hits in order to test the likeliness that they could be in accordance with the proposed formula. This means that if the fragment formula would contain S, but none of the molecular formulae do, this possibility can be rejected. Components that showed a library match < 700, *e.g.* those that were probably not registered in our NIST library, were selected for elucidation in the course of this discussion.

All the samples were re-injected into the GC-MS system using the CI source in positive and negative mode. These data were used to identify the intact molecule. Once the intact molecule was identified from the GC-CI-MS data, the accurate mass for the protonated molecule was submitted to the calculation of all possible formulae with a maximum deviation of 5 mDa from the measured mass using the Elemental Composition Calculator. Parameter settings for all calculations were C: 0–30, H: 0–50, N: 0–10, O: 0–10, and P: 0–3. It is worth to notice that other authors have deeply studied these constraints using the development set of formulas derived from NIST and Wiley and finally proposed a maximum element count defined for different mass ranges [10]. In principle, no F atoms were considered, as this would considerably

increases the number of possible elemental compositions, which would complicate the elucidation. However, if evidence on the presence of F atoms in the molecule were observed in the experimental MS data from the loss of F• or HF, the presence of F (0–10) was obviously considered during the elucidation step. In addition, from the characteristic isotopic patterns associated to ^{37}Cl (31.98% relative abundance), ^{81}Br (97.88%) and ^{34}S (4.44%), the appropriate number of Cl, Br and S atoms was evaluated and added. The number of Cl and Br atoms was easily adjusted. However, the lower relative abundance of ^{34}S made the adjustment of S atoms less precise, especially when halogens were also present. In these cases, an interval was given. The accepted deviations between the experimental and the theoretical values were empirically derived in accordance with previous work in our research group. Briefly, when the abundance of an isotopic peak was between 60 and 200 counts, the observed accepted deviation was 20%, and when the abundance was higher than 200 counts, the error decreased to below 10%.[16]

A carbon number prediction filter of ±5 was applied to reduce the number of possible elemental compositions for a particular mass if the intensity of the molecular ion in the spectrum was higher than 300 counts. The double-bond equivalent (DBE) parameter was set from -1.5 to 50, but was not used as an identification criterion, although information about aromaticity of the structure was obtained. Additionally, the option ‘‘even-electrons ions only’’ was selected for the (de)protonated molecule in CI ionization data. Fragment ions present in the CI spectrum were used to enable a

further reduction in the number of possible molecular formulae; the option “odd and even-electron ions” was used for this purpose. Also, accurate mass data on EI fragment ions were used to reduce the number of possible molecular formulae, e.g., because particular fragment ions cannot be generated from a particular molecular formula (examples: see below). The option “odd-electron ions only” was selected for the molecular ion in EI data (if it existed) and ‘‘odd- and even-electrons ions’’ was used for the fragment ions. Similarly, a carbon number prediction filter of ±5 was applied to reduce the number of possible formulae in the spectrum if the intensity of the ion in the spectrum was higher than 300 counts. It is worth to notice that, in the case of fragment ions, the carbon filter should applied with care as an additional McLafferty rearrangement might occur during the fragmentation process and (apparently) disturb the expected isotopic pattern. Once a formula was elucidated, it was searched in databases. We have chosen the Reaxys database (www.reaxys.com), a web-based search and retrieval system for chemical compounds, bibliographic data and chemical reactions that contains more than 18.000.000 substances. In some cases, EI spectra provide valuable information about a substructure of the unknown. Reaxys allows limiting the search of a formula taking into account a substructure. This notably reduces the number of possible structures for a given formula. For the structures finally proposed, the fragmentation patterns observed in the EI and CI spectra could be explained. Although not applied in this paper, other tools, as prediction of retention times on the used GC-column, might be helpful to reduce the number of candidates delivered by databases.

RESULTS AND DISCUSSION

Accurate masses alone do not allow the retrieval of correct elemental formulae due to the large search space of chemically possible solutions. So, a combination of different rules that constrains and scores all chemically possible formulae based on accurate mass measurements, the formulae proposed, their isotopic patterns, carbon number prediction filter and nitrogen rule, among other, are necessary.

In the process of chemical identification of unknown compounds, it is important to obtain overall high signal intensities for molecular ions (or defined adducts or fragments of molecular ions) and therefore optimal signal-to-noise ratios for each peak. Higher signal intensities yield better ion statistics, thus improving accurate mass and isotopic abundance measurements, which subsequently lead to higher confidence in determining elemental compositions. Electrophilic addition in positive-ion CI fairly often gives rise to $[M+C_2H_5]^+$ and $[M+C_3H_5]^+$ adduct ions next to $[M+H]^+$. Thus, [M+29.0391] and [M+41.0391] peaks may be observed in addition to the expected [M+1.0078].

In this work, a maximum deviation of 5 mDa in measured masses was selected for elemental composition calculation. This value may seem a bit high considering the capabilities of modern mass spectrometers. Thus, we have chosen a less favorable scenario to fully explore the potential of our TOF MS under more critical situations. The choice of 5 mDa mass errors has been made according to our own experience in analysis of known compounds with our GC-TOF MS. Obviously, lowering the mass error tolerance would decrease the number of potential candidates for a given mass, but it would also increase the possibilities of losing the right elemental composition if the instrument for whatever reason does not satisfy the strict requirements established.

Carbon filtering

The Elemental Composition Calculator within the MassLynx software allows calculating possible formulae using predefined parameter settings. Among these parameters, the element prediction filter applied to estimate the number of carbons of the unknown structure reduces considerably the number of suggested formulae returned by the program. For this purpose, a carbon range must be defined by the user to exclude all suggestions that fall outside an estimated range of carbon atoms for the molecule of interest. The number of carbon atoms in a molecule can be estimated by considering the relative intensity of the "M+1" isotope peak which, in the absence of Si, is mainly due to the presence of $^{13}C_1$. With the carbon filtering, the Calculator returns only those results that include the estimated number of carbons, plus or minus the number of carbons entered by the user. An incorrect use of this option can unwittingly exclude the correct composition if the experimental data does not correctly reflect the mass and isotope pattern of the compound. In order to asses the most appropriate carbon range to be applied, a systematic study was carried out on the error in the estimation of the number of carbons in a molecule from the "M+1" peak using a series of standards in solvent. A mixture of several pesticides (dichlorvos, lindane, diazinon, chlorpyriphos methyl, pirimiphos-methyl, fenthion, simazine, terbutylazine, diphenylamine and molinate) at different concentration levels were injected into the GC-MS system under positive ion CI conditions. Considering the relative intensity of "M+1" isotope peaks, which under these conditions is the peak with *m/z* of $[M+H]^+$+1, the experimental number of carbons of the molecule was estimated and compared with real value for the target molecule. Experimental results showed that when the peak intensity was below 300 counts, no M+1 could be observed in the spectrum. In such cases, no carbon filter could be applied. When the intensity of the peak was higher than 300 counts, the estimated number of carbons generally did not differ by

3 or 4 from the true value. This led us to conclude that a carbon filter of ±5 would be a good choice (for peaks with intensities higher than 300 counts).

Selected examples

In this paper, we show selected examples for the elucidation of model compounds, which have been chosen because their mass spectra are not registered in the commercial library available in our laboratory NIST02 library. The pesticides discussed here as illustrative examples are bifenazate, boscalid, epoxiconazole, and fenhexamid. A ground water extract was spiked with a mixture of these pesticides and injected into the GC-TOF MS, under EI and CI conditions. Then, TOF MS full-acquisition data were processed treating the sample as unknown, i.e. a non-target processing method was applied without using any previous information on analyte identity.

Case 1

From the EI data and applying a non-target screening approach in the ChromaLynx software, a chromatographic peak with a retention time 31.6 min was found that returned a match of 670 in the NIST library search, indicating the compound is 1,1'-biphenyl, 4-methoxy (M 184,0875 Da). The accurate mass of the protonated molecule of this unknown compound was determined to be *m/z* 301.1563 from the CI+ spectrum (Figure 1A), indicating the match from the library search is not correct. Within the search limits outlined above, calculation of the possible elemental compositions resulted in 13 formulae. When applying the carbon filter, 5 formulae remained (Figure 1B). A fragment ion with *m/z* 259.1068 present in the positive CI spectrum corresponds to the loss of 42.0495 Da that could be due to the loss of C_3H_6 (42.0470).

Looking at the EI spectrum (Figure 1A, top), the major fragment is an ion with *m/z* 184.0881.

For this *m/z*, 13 possible formulae are obtained, which number reduces to only 3, if the carbon filter is applied (Figure 1B). Other fragment ions present in the EI TOF MS spectrum could be considered as subsequent losses of $CH_3^{\bullet}$ (*m/z* 169.06578), CO (*m/z* 141.0708) and C_2H_2 (*m/z* 115.0551), which allowed us to discard 1 out of 3 formulae for the ion with *m/z* 184.0881 (the one without O) remaining $C_{13}H_{12}O$ or $C_9H_{15}NOP$. These two formulae allowed us to discard 1 out of 5 initial formulae calculated from the unknown protonated molecule, remaining $C_{13}H_{17}N_8O^+$, $C_{15}H_{26}O_4P^+$, $C_{17}H_{21}N_2O_3^+$, and $C_{14}H_{27}N_2OP_2^+$.

In this particular case, the EI spectrum possibly provides substructure information on the unknown, based on *m/z* 184 and its three fragment ions. The unknown would (most likely) contain a methoxy-substituted biphenyl, that is "for instance" 1,1'-biphenyl-4-methoxy. The methoxy-group could be at another position. By a search through the Wiley EI-MS Library, it appears both 2- and 4-methoxy-1,1'-biphenyl give about the same spectrum. This allowed us to discard again 2 out of 4 remaining formulae, as the unknown should have a DBE of at least 8 (due to the biphenyl group). At this point, still two formulae remained, $C_{17}H_{21}N_2O_3^+$ (DBE = 8.5) and $C_{13}H_{17}N_8O^+$ (DBE = 9.5). Looking at the positive CI spectrum (Figure 1A, bottom), subsequent losses of C_3H_6 and CH_3NO_2 can be observed which only can be accomplished from $C_{17}H_{21}N_2O_3^+$.

A Reaxys database search was performed and the elemental composition $C_{17}H_{20}N_2O_3$ resulted in 1492 structures. Limiting the above search to the substructure revealed by EI spectrum (methoxy-substituted biphenyl), a total of only five structures were returned by the database (Figure 1C).

As commented above, in the CI spectrum we observed the loss of 42.0472 Da (C_3H_6) to an ion with *m/z* 259.1107. This loss allowed us to discard structures 2, 3 and 5.

From structure 1, both the formation of the odd-electron fragment ion with *m/z* 184 in EI and the fragment ions with *m/z* 259 and 198 are readily explained (see Figure 2) whereas the odd-

electron fragment ion with *m/z* 184 in EI and the even-electron fragment ion with *m/z* 198 in CI are not expected to be formed from structure 4. In this case, in order to generate an ion with *m/z* 184, two different bonds to the ring should be cleaved. The NO_2-group would be lost as a radical (prior to or after the loss of propylene (C_3H_6). From the resulting even-electron structure, it would be highly unlikely to lose the other side chain in such away that an ion with *m/z* 184 is formed. In CI, the formation of the ion with *m/z* 198 would require the loss of two radicals: NO_2 and CH_3. Consequently, in the light of the results obtained, we proposed the structure 1 for this compound, which in fact corresponds to acaricide bifenazate, the pesticide already present in the water sample.

Case 2

The second example involves a compound with a retention time of 36.6 min and a protonated molecule with *m/z* 343.0443 (Figure 3, middle). Typical adducts with $C_2H_5^+$ and $C_3H_5^+$, consequence of the use of methane as a reagent gas in CI mode, were observed in the positive CI spectrum. The experimental M+2 abundance of 69.6%indicates the presence of two Cl atoms in the molecule, and nil to two S atoms, as deduced from the accepted tolerances in the M+2 percentage (±10%, i.e., 62.6-76.6%). Within the procedure outlined above, only one formula remained ($C_{18}H_{13}Cl_2N_2O^+$). A Reaxys database search was resulted in 81 structures. Based on library matches, the EI spectrum provides useful substructure information (Figure 3, bottom). The unknown compound most likely is an ester/amide of monochloro-pyridine carboxylic acid, with an additional Cl in the other part of the molecule. Limiting the above search with this substructure ($C_6H_3ClNO^+$), only one structure is returned by the database corresponding to the pyridinecarboxamide fungicide boscalid ($[M+H]^+$ with *m/z* 343.0405). However, it is difficult to propose structures for the poor-abundance high-*m/z* fragments from this structure (loss of water, followed by loss of $Cl^•$, followed by loss of HCl to *m/z* 253) (see

figure 3, bottom). Obviously, the most logical next step would be to purchase the reference standard of the suggested candidate, boscalid, check its retention time and mass spectra in the various modes for definitive confirmation.

Case 3

The accurate mass of a protonated unknown compound with retention time 30.67 min was determined to be *m/z* 330.0820; characteristic adducts with $C_2H_5^+$ and $C_3H_5^+$ were also observed in the CI+ spectrum (Figure 4A, middle). Given the experimental M+2 abundance of 34.2 %, we assumed the presence of one Cl atom in the molecule, and nil to one S atoms, as deduced from the accepted tolerances in the M+2 percentage (±10%, i.e., 30.8-37.6%). Calculation of the possible formulae yielded 5 results. The fragment ion with *m/z* 310.0743 present in the positive CI spectrum corresponded to the loss of 20.0077 Da that can only be due to the loss of HF (20.0062 Da). At that point, the possible presence of 1-10 F atoms was included in the calculations. Within the new limits, the calculation resulted in 7 possible formulae (Figure 4B). The EI fragment ion with *m/z* 313.0758 corresponded to the loss of 15.9937 Da that only can be the loss of O (15.9949 Da) (Figure 4A, bottom). This oxygen loss allowed us to discard the 3 molecular formulae that did not contain any atom of oxygen. At this point, 4 molecular formulae still remained. From calculation of the formula of the EI fragment ion with *m/z* 244.0457, only one possible formula was found ($C_{15}H_{10}FCl^{+\bullet}$) (Figure 4B). Using this information, two formulae containing less than 15 carbon atoms could be discarded from our list. The two remaining formulae were $C_{15}H_{19}ClFNO_2P^+$ and $C_{17}H_{14}ClFN_3O^+$ (protonated molecules). The first formula ($C_{15}H_{19}ClFNO_2P^+$) is highly unlikely, as the generation of the fragment ion with *m/z* 244 would require the loss of $H_8NO_2P^\bullet$. The rest of the CI and EI fragments did not help us to discard any of the two empirical formulae. So, a Reaxys database search was performed. The formula

$C_{15}H_{18}ClFNO_2P$ did not result in a structure. However, the formula $C_{17}H_{13}ClFN_3O$ resulted in 44 possible structures in Reaxys, among which there are a number of stereoisomers that cannot be differentiated by MS. At this point, a substructure is needed to reduce the number of possibilities. A possible substructure may be derived from further interpretation of CI spectrum. Next to the loss of HF, the loss of 69.0305 Da is observed, which could be consistent with $C_2H_3N_3$ (69.0327 Da), a triazole substructure. The complementary fragment with *m/z* 70 is also observed (Figure 4A, middle). In fact, the fragment ion with *m/z* 244 in the EI spectrum is due to a combined loss of oxygen and the triazole ring. This triazole substructure is found in thirteen of the 44 possible structures found in the Reaxys database (Figure 4C). However, among these thirteen, there are 9 stereoisomers of the same structure (structure 1). From three other structures, an easy loss of the triazole ring, as observed in both the EI and the CI spectrum, is not likely either because it requires the cleavage of too many bonds or a massive rearrangement (structures 2, 3 and 5). This means that only two isomeric structures are left (1 and 4). A choice between these two can (possibly) only be made from differences in retention time. Both these structure proposals enables us to explain the fragments in the negative-ion CI spectrum: the loss of $Cl^{\bullet}$ leads to the fragment ion with *m/z* 293, the loss of $C_3H_3N_3^{\bullet}$ to *m/z* 247, and the loss of $Cl^{\bullet}$ and $C_2H_3N_3$ to *m/z* 224.

At this stage, it must be admitted that this particular case also indicates one of the weak points of the current procedure. The recognition of relevant substructures seems to be an issue of experience and a bit of luck. In this case, the EI fragment ion with *m/z* 139 ($C_7H_4ClO^+$, that is most likely Cl–phenyl–C≡O^+) was considered as a relevant substructure (Figure 4A, bottom). This would indicate that the unknown most likely is an ester/amide of chlorobenzoic acid. Performing a Reaxys database search with $C_{17}H_{13}ClFN_3O$ and this substructure returned only one possible structure, from which the loss of the triazole ($C_2H_3N_3$) is not likely. Although not recognized by us, the formation of the Cl–phenyl–C≡O^+ fragment apparently is possible from

the epoxide structure proposed.

Interestingly, this apparently possible substructure with *m/z* 139 allowed us to provisionally differentiate between the two possible structures left (1 and 4, see above). From structure 1, the formation of Cl–phenyl–C≡O^+ is readily expected (Figure 5), whereas with structure 4, the formation of F–phenyl– C≡O^+ would be more likely. Consequently, we propose the structure 1 for this compound, which in fact corresponds to epoxiconazole, the pesticide already present in the water sample.

Case 4

In this case, a chromatographic peak was detected in the non-target screening with retention time 29.45 min that returned a match of 606 in the NIST library search, indicating the compound to be 1-methyl-cyclohexene. The accurate *m/z* of the protonated molecule of this unknown compound was *m/z* 302.0732 (Figure 6 middle). Within the search limits explained above, calculation of the possible elemental formulae resulted in only 1 formula ($C_{14}H_{18}Cl_2NO_2^+$). This formula is consistent with a DBE of 6. A Reaxys database search resulted in 171 possible structures. The positive-ion CI spectrum shows little fragmentation, whereas in negative-ion CI spectrum, only a loss of a $Cl^\bullet$ is observed (Figure 6a top). The data from the library search are rather non-informative, except that the possible presence of a methyl-substituted cyclohexane substructure is suggested. This is somewhat confirmed by the loss of C_7H_{14} from the molecular ion (*m/z* 301 to 203). The calculated formula for the resulting fragment ion with *m/z* 203 is $C_7H_3Cl_2NO_2^{+\bullet}$ (DBE=6), indicating most likely the presence of a dichloro-substituted benzene or pyridine ring next to the methyl-substituted cyclohexane (Figure 6a bottom). Three separate database searches were performed (one for each substructure). From the search results, those structures were selected, which showed both substructures, as both the aromatic and the non-aromatic rings are part of the structure.

After that, only two isomeric structures remained (Figure 6b) with the weak bond in the ester or amide link. Any MS fragmentation will lead to a 4-amino-2,3-dichlorophenol (*m/z* 177) and/or 2,3-dichloro-4-iminocyclohexa-2,5-dien-1-one (*m/z* 175) type of fragment (*m/z* 175, 177, 179), from which one never could be decided how this part is attached to the remainder of the molecule, that is via O or N. So, at this point, the only way to discriminate between the two is by checking the retention time after the injection of references standards. Structure (1) in Figure 6b is fenhexamide, the pesticide already present in the water sample.

CONCLUSIONS

In this paper, the complementary use of EI and CI has been investigated in combination with GC-TOF MS for the elucidation of organic (micro)contaminants in water samples. Several model examples have been shown to illustrate the methodology applied and the difficulties of this process when the mass spectra of the compounds investigated are not available in the commercial library used in our laboratory. The use of the soft-ionization technique CI, has allowed the determination of the molecular mass. In addition, accurate mass measurement provided by TOF MS, together with the structure information generated by the accurate mass EI spectrum, has allowed the proposal of an appropriate formula for the unknown. The application of rules based on observed isotopic patterns, carbon number prediction filter and nitrogen rule, among others, has been crucial to reduce the number of possible formulae. Searching the candidate formulae in a database has allowed the proposal of chemical structures for the unknown. The recognition of relevant substructures on the unknown molecule has been of great help in order to reduce the number of possible structures given by the database search. At this stage, the recognition of relevant substructures seems to be an issue of experience, which reflects the difficulties of this challenging task. Accurate

masses of fragment ions given by TOF MS are of outstanding importance. Their structures should be compatible with the chemical structure assigned to the candidate. In this work, the unknown compound could be identified in several cases, while in others two chemical possible (isomeric) structures remained as candidates. At this point, the unequivocal confirmation should be made by injecting the reference standard, if available, to test the retention time and experimentally confirm the presence of fragment ions generated by GC-TOF MS. According to our experience, elucidation of unknowns cannot be easily made following a completely standardized procedure, as both expertise and creativity are fully necessary in such a process. In this paper, we applied a general methodology, but found out and demonstrated that this methodology has to be somewhat adapted depending on the spectral information and the results of the database search.

ACKOWLEDGMENTS

The authors are very grateful to the Serveis Centrals d'Instrumentació Científica (SCIC) of University Jaume I for the use of GC-TOF MS instrument. This work has received financial support from the Spanish Ministry of Science and Education (Project reference: CTQ2009-12347). The authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence PROMETEO/2009/054. Tania Portolés is very grateful to University Jaume I for her post-doctoral grant.

FIGURE CAPTIONS

Figure 1. (A) Electron ionization (top) and positive chemical ionization (bottom) spectra of a detected compound. (B) Possible elemental compositions for different *m/z* ions after applying carbon filtering. (C) Possible structures for $C_{17}H_{20}N_2O_3$ limiting the Reaxys search to the methoxy-substuted biphenyl substructure.

Figure 2. Structures suggested for different *m/z* ions taking into account structure 1 (bifenazate). Electron ionization (top) and positive chemical ionization (bottom) spectra.

Figure 3. Structures suggested for different *m/z* ions taking into account the boscalid structure. Negative ion chemical ionization (top), positive ion chemical ionization (middle) and electron ionization (bottom) spectra.

Figure 4. (A) Negative ion chemical ionization (top), positive ion chemical ionization (middle) and electron ionization (bottom) spectra of a detected compound. (B) Possible elemental compositions for different m/z ions after applying carbon filtering. (C) Possible structures for $C_{17}H_{13}ClFN_3O$ limiting the Reaxys search to the triazole substructure.

Figure 5. Structures suggested for different *m/z* ions taking into account the structure of epoxiconazole. Negative ion chemical ionization (top), positive ion chemical ionization (middle) and electron ionization (bottom) spectra

Figure 6. (a) Negative ion chemical ionization (top), positive ion chemical ionization (middle) and electron ionization (bottom) spectra of a detected compound. (b) Possible structures for $C_{14}H_{18}Cl_2NO_2$ limiting the Reaxys search to different substructures

REFERENCES

1. Hernández F, Portolés T, Pitarch E, López FJ. *TrAC-Trends in Anal. Chem*. 2010, DOI: 10.1016/j.trac.2010.11.007.
2. Abate S, Ahn YG, Kind T, Cataldi TRI, Fiehn O. *Rapid Commun. Mass Spectrom.* 2010; **24**: 1172.
3. Grange AH, Genicola FA, Sovocool GW. *Rapid Commun. Mass Spectrom.* 2002; **16**: 2356.
4. Carrasco-Pancorbo A, Nevedomskaya E, Arthen-Engeland T, Zey T, Zurek G, Baessmann C, Deelder AM, Mayboroda OA. *Anal. Chem.* 2009; **81**: 10071.
5. Hejazi L, Ebrahimi D, Hibbert DB, Guilhaus M. *Rapid Commun. Mass Spectrom.* 2009; **23**: 2181.
6. Schymanski EL, Bataineh M, Goss K-, Brack W. *TrAC - Trends in Anal. Chem.* 2009; **28**: 550.
7. Schymanski EL, Meringer M, Brack W. *Anal. Chem.* 2009; **81**: 3608.
8. Schymanski EL, Meinert C, Meringer M, Brack W. *Anal. Chim. Acta* 2008; **615**: 136.
9. Schulze T, Weiss S, Schymanski E, von der Ohe PC, Schmitt-Jansen M, Altenburger R, Streck G, Brack W. *Environ. Pollut.* 2010; **158**: 1461.
10. Kind T, Fiehn O. *BMC Bioinformatics* 2007; **8**. art 105.
11. Kind T, Fiehn O. *Bioanal. Rev.* 2010; **2**: 23.
12. Fay LB, Newton A, Simian H, Robert F, Douce D, Hancock P, Green M, Blank I. *J. Agric. Food Chem.* 2003; **51**: 2708.
13. Wanying Y, Hua Z, Weidong H, Jiping C, Xinmiao L. *Talanta* 2005; **65**: 172.
14. Ojanperä S, Pelander A, Pelzing M, Krebs I, Vuori E, Ojanperä I. *Rapid Commun. Mass Spectrom.* 2006; **20**: 1161.
15. Tyrkkö E, Pelander A, Ojanperä I. *Drug Testing and Analysis* 2010; **2**: 259.

16. Ibáñez M, Sancho JV, Pozo OJ, Niessen W, Hernández F. Rapid Commun. Mass Spectrom. 2005; 19: 169.
17. Hernández F, Portolés T, Pitarch E, López FJ. Anal. Chem. 2007; 79: 9494.
18. Portóles T, Ibáñez M, Sancho JV, López FJ, Hernández F. J. Agric. Food Chem. 2009; 57: 4079.
19. Hernández F, Portolés T, Pitarch E, López FJ. J. Mass Spectrom. 2009; 44: 1.