Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1067 (2005) 127–134
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Matching unknown empirical formulas to chemical structure using LC/MS TOF accurate mass and database searching: example of unknown pesticides on tomato skins

E. Michael Thurman*, Imma Ferrer, Amadeo Rodriguez Fernández-Alba

Pesticide Residue Research Group, University of Almería, 04120 Almería, Spain

Available online 23 December 2004

Abstract

Traditionally, the screening of unknown pesticides in food has been accomplished by GC/MS methods using conventional library searching routines. However, many of the new polar and thermally labile pesticides and their degradates are more readily and easily analyzed by LC/MS methods and no searchable libraries currently exist (with the exception of some user libraries, which are limited). Therefore, there is a need for LC/MS approaches to detect unknown non-target pesticides in food. This report develops an identification scheme using a combination of LC/MS time-of-flight (accurate mass) and LC/MS ion trap MS (MS/MS) with searching of empirical formulas generated through accurate mass and a ChemIndex database or Merck Index database. The approach is different than conventional library searching of fragment ions. The concept here consists of four parts. First is the initial detection of a possible unknown pesticide in actual market-place vegetable extracts (tomato skins) using accurate mass and generating empirical formulas. Second is searching either the Merck Index database on CD (10,000 compounds) or the ChemIndex (77,000 compounds) for possible structures. Third is MS/MS of the unknown pesticide in the tomato-skin extract followed by fragment ion identification using chemical drawing software and comparison with accurate-mass ion fragments. Fourth is the verification with authentic standards, if available. Three examples of unknown, non-target pesticides are shown using a tomato-skin extract from an actual market place sample. Limitations of the approach are discussed including the use of A + 2 isotope signatures, extended databases, lack of authentic standards, and natural product unknowns in food extracts.

Keywords: Pesticides; Time-of-flight; Mass spectrometry; Tomato-skin extract; Library; Database

1. Introduction

The identification and quantitation of unknown pesticides in vegetables is of great importance to individuals and health organizations around the world. In order to meet these health concerns, the European Union (EU) and the US have set new directives for pesticides at low levels in vegetables. For example, new laws, such as the European Directive 91/414/EEC or the Food Quality Protection Act (FQPA) in the US, have changed the standards for human health, workers, and environmental protection, which require lower levels of pesticides in food [1–2]. They also require re-registration for older pesticides [1–2]. Furthermore, the review programs have withdrawn authorizations for many of the crop protection products currently on the market, 177 compounds in US and 320 in Europe. Moreover, it was announced in Europe that a total of 110 products would be withdrawn in the near future (EU Directive 91/414/EEC, reference [1]).

Next, the quality standards within the new regulations include the re-assessment of the maximum residue limits (MRLs) for vegetables. The EU directives are setting different MRLs for each pesticide within each food group, and typically, the MRLs are lower than the previous ones. Furthermore, the new directives also lead to different MRLs for each EU country, which are still being decided. The EU directives

* Corresponding author.

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.11.007

state that the individual country MRLs will be maintained in the new program. Finally, banned compounds have the lowest MRLs, which is set now at 0.01 mg/kg (ppm). With the planned program to remove so many compounds from the market, it is important, even necessary, that screening for unknown pesticides may be done by both GC/MS and LC/MS on vegetable extracts. Because it is not always possible to know which banned substances may be used, it is of vital importance to environmental food monitoring that there be a system of unknown identification to give fast and accurate screening of unknown substances in food and food products by both GC/MS and LC/MS methods, which are complementary techniques. This paper focuses on new advances in accurate mass LC/TOF/MS for the identification of unknown, non-target pesticides on vegetables, in particular routine sub-2 ppm mass accuracy.

Thus, there is an important need for research studies and methods development on the analysis of unknown non-target pesticides in vegetables by new LC/MS methods, such as the combination of accurate mass using LC/MS TOF and MS/MS using LC/MS ion trap and LC/MS/MS in general [3–8]. Our study in this report is one of the first of its kind to examine LC/MS TOF combined with LC/MS ion trap, and the use of commercial databases, such as the Merck Index and the ChemIndex database to identify unknown pesticides in food. This comment is based on the recent review (2004) of LC/MS analysis of pesticides in food by Pico et al. [9], which include no LC/TOF/MS papers.

Several advantages of the combination of LC/MS TOF and ion trap are that accurate mass and empirical formulas may be combined with the MS/MS spectra and MS^n for spectral information [3–6,10–11]. The use of LC/MS Q/TOF has also been a successful technique for unknown identification [3,11], although it lacks the sensitivity of the ion trap in full scan mode and is not capable of MS^n, which is sometimes valuable in unknown identification.

The concept of unknown identification using the LC/MS TOF and ion trap consists of four steps, which are outlined in detail below. They are:

1. analyze the vegetable extract with LC/MS TOF in full scan looking for large unknown peaks using a mild in-source CID fragmentation typically with positive ion electrospray (examples here show positive ion only as there were no large detected peaks in negative ion only background).
2. Search Merck Index or ChemIndex for unknowns using the generated empirical formulas and any A + 2 isotopes, such as Cl, Br, or S, if present.
3. Proceed to ion trap MS/MS with proposed structures and do MS^2 or MS^3. Use a chemical-structure drawing program to identify ion fragments and their accurate masses. Then, combine with LC/MSD TOF data of fragment ions (empirical formula of fragment ion), if available. Make tentative identification.
4. Obtain and analyze standard for final confirmation, if available. This report gives three detailed examples of this process using store purchased tomatoes, which contained "unknown white powders" that were subsequently identified by the above process for various "unknown, non-target pesticides" on the skin.

2. Experimental methods

2.1. Vegetable-skin extraction

Selected tomatoes containing white powder from a commercial market place were extracted as follows. Carefully wash the skin of the tomato three times with methanol to remove the white powder, from 2 to 5 mL, depending on the size of the vegetable. Capture the solvent in a 150 mL Pyrex beaker. After mixing, transfer the methanol to a 5 mL syringe and filter through a Millex®-FH PTFE filter and aliquot 0.3 mL. Dilute with 0.6 mL of de-ionized water. Analyze by either LC/MSD TOF or LC/MSD ion trap directly.

2.2. LC/MS TOF methods

LC Pumps were HP 1100, injection volume 50 μL, column: ZORBAX Eclipse® XDB 4.6 mm × 150 mm C-8.5 μm, mobile phase A = ACN and B = 0.1% formic acid in water, gradient was 15–100% A over 30 min at a flow rate of 0.6 mL/min, model LC/MSD TOF (Agilent Corp, Santa Clara, CA, USA) with electrospray source positive ESI+, capillary 4000 V, nebulizer 40 psig, drying gas 9 L/min, gas temp 300 °C, fragmentor 190 V, skimmer 60 V, Oct DC137.5 V, OCT RF V 250 V, reference masses: 121.0509 and 922.0098 *m/z*, resolution: 9500 ± 500 @ 922.0098 *m/z*. Reference A sprayer 2 is constant flow rate (100 μL/min) during the run. Reference masses consist of fluorinated unknown compounds furnished by the manufacturer with empirical formulas. Formula calculator included the atoms: C = 50, H = 100, N = 10, O = 10, P = 1, S = 2, Cl = 3, and F = 5. Accuracy checks of the instrument were carried out with the LC/TOF/MS analysis of atrazine (accurate mass *m/z* 216.1010) within 2 ppm prior to instrument operation and use.

2.3. LC/MS ion trap methods

LC Pumps were HP 1100, injection volume 50 μL, column: ZORBAX Eclipse® XDB 4.6 mm × 150 mm C-8, 5 μm, mobile phase A = ACN and B = 0.1% formic acid in water, gradient was 15–100% A over 30 min at a flow rate of 0.6 mL/min, Model LC/MSD Trap (Agilent, Santa Clara, CA, USA) with electrospray source positive ESI+, capillary 3200 V, nebulizer 40 psig, drying gas 9 L/min, gas temp 300 °C, fragmentor 70 V.

2.4. Database and chemical drawing software

The databases searched included two CD ROM databases, The Merck Index and ChemIndex, both of which are com-

mercially available from CambridgeSoft in Cambridge, Massachusetts (USA), and not part of the software package of the LC/TOF/MS instrument. Likewise, the chemical drawing software was ChemDraw also from CambridgeSoft. The chemical drawing software has the capability to do accurate mass analysis for either GC/MS (electron impact) or LC/MS (electrospray), including adducts and protonated molecules. The chemical drawing software is not part of the LC/TOF/MS instrument software. The database and chemical drawing software are sold as a package called ChemOffice (CambridgeSoft, Cambridge, MA, USA) and run on a windows environment (e.g. laptop). Likewise, the ion trap (beta version) and LC/TOF/MS software are available from Agilent for data analysis on a laptop windows environment. This is a useful combination for rapid work on unknown identification not requiring the instrument or its computer system.

3. Results and discussion

3.1. Tomato-skin extract

Fig. 1 shows the total ion chromatogram (TIC) for the rapid extraction of a white powder on a store-purchased tomato (skin) using LC/TOF/MS. The simplicity of the extraction of the tomato skin results in a clean chromatogram without the interferences of the much of the matrix of the tomato. Several recent studies have shown that the skins of vegetables contain high concentrations of pesticides [12–13]; thus, this extract is a good medium for unknown pesticide identification. Furthermore, it is an environmentally relevant extract since the skins of tomatoes are eaten in salads and extracted for many food uses.

The LC/TOF/MS instrument of this study (Agilent) is one of the first of its type to use a analog-to-digital converter (ADC) instead of a digital-to-time converter (TDC) for the taking and averaging of mass spectral peaks. Discussions of ADC to TDC state generally that there is a wider window of sample intensity and, by inference, mass accuracy across a wider concentration range, before saturation of the detector with the ADC type. The ADC detector was used in this study and it was found that slicing the peak was of no advantage over taking the entire peak for mass accuracy. Thus, the method used here was to take the center 95% of the peak for mass-accuracy measurements up to an intensity equal to the calibration ions (intensity of 200,000 counts). If intensities exceeded the calibration ions by greater than 50% the accuracy measurement was taken off center to reduce counts

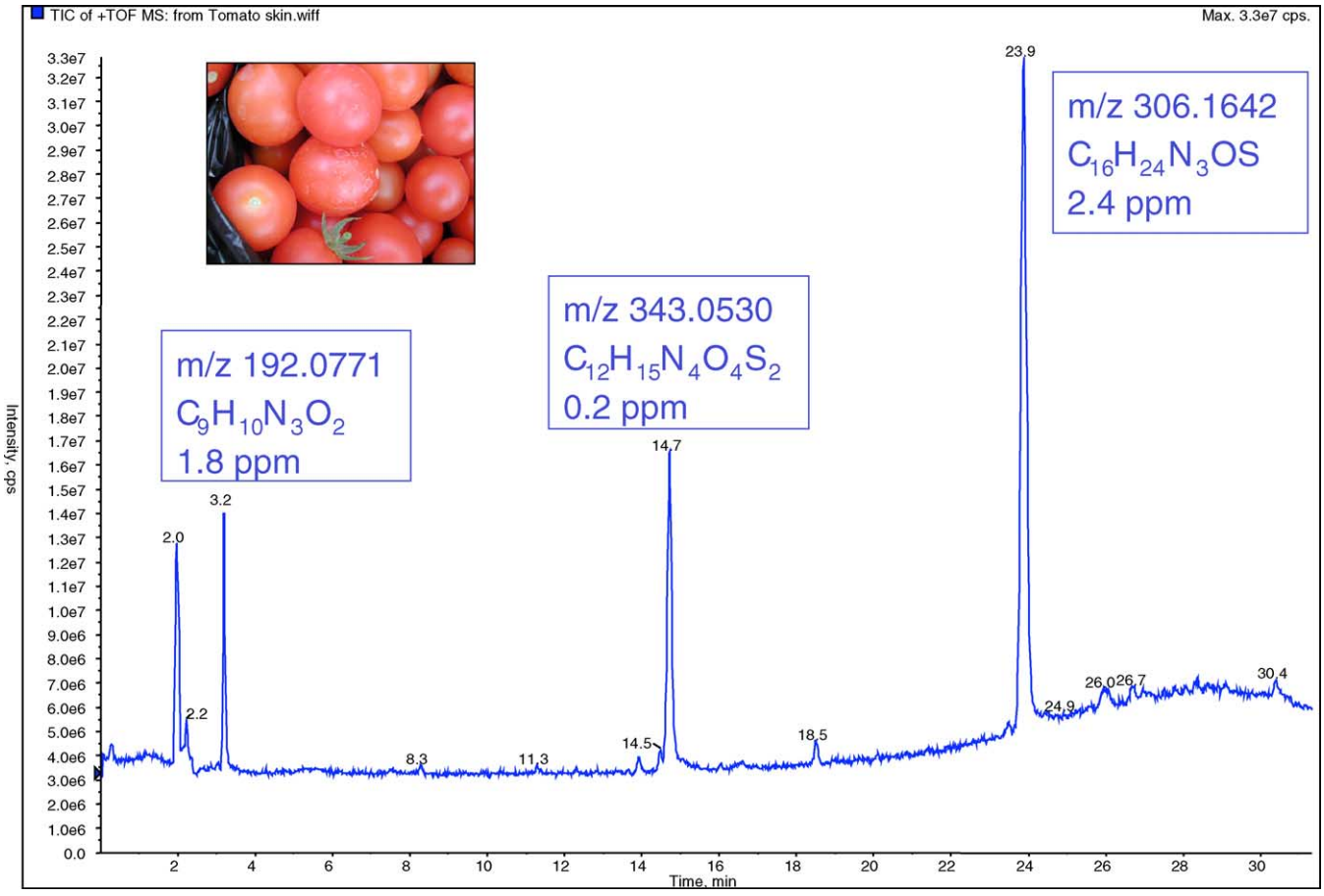

Fig. 1. LC/MS TOF with accurate mass, formulas, and ppm error of three major peaks in TIC of tomato-skin extract.

or the sample was diluted and re-analyzed. We did notice a deviation from the 3 ppm accuracy limit of the manufacturer if the intensity of the unknown ion was greater than 5–10 times the intensity of the calibration ion. This does indicate that saturation of even the ADC detector can occur with large peak intensities. The remedy is to dilute the sample at least ten times and re-analyze. Intensities as much as 10 times less than the calibration ion may also cause deviation from the 3 ppm specifications and it may be important to concentrate the sample before analysis for the highest accuracies (<3 ppm).

The chromatogram (Fig. 1) is much simpler than a whole tomato extract because of the lack of natural product peaks, while remaining effective on pesticides on the surface of the tomato. For example, there are four major peaks in the chromatogram at retention times of 2.0, 3.2, 14.7, and 23.9 min. The peak at 2.0 min is the void volume of the LC column. For purposes of discussion of the data, let us begin with the peak at 14.7 min in the TIC. The accurate mass *m*/*z* is 343.0530 with an A + 2 isotope of *m*/*z* 345.0499 that is 9.8% of the main peak (Fig. 2). This percentage of 9.8% suggests that an A + 2 isotope of S-34 is present with two atoms (4.2% for each S-34 is the natural distribution). Four possible formulas were found at an error of <3 ppm using the calculator tool of the software (Fig. 2) and forcing the formulas with 1S atom (without forcing the S atom there were 35 hits using atoms listed in Experimental section and the manufacturer's accuracy limit of 3 ppm). However, only the first formula contains two sulfurs and using the isotope matching tool of the software, this formula was the only one that gave a perfect match (Fig. 2, notice the dotted lines, which indicate the matching). It is important to note that even with an accuracy of 3 ppm it is possible to have many formulas to possibly search; thus, the use of the A + 2 isotope is quite helpful in limiting the list of targets for searching.

Thus, this formula (and the three others) were entered into the Merck data base for searching and no formula matches were found. The search was repeated with the ChemIndex data base (Cambridge Software) and the formula and structure of thiophanate methyl was located (Fig. 3). Note that a hydrogen atom was removed from the structure before the search because the ion at *m*/*z* 343.0530 contains an extra proton that is not present in the database empirical formula. This compound is used as a fungicide on fruits and vegetables [14] in order to protect from mold.

The next step was to analyze the sample by LC/MS ion trap and look for fragment ions that would result from the fragmentation of thiophanate methyl. Fig. 4 shows the ion

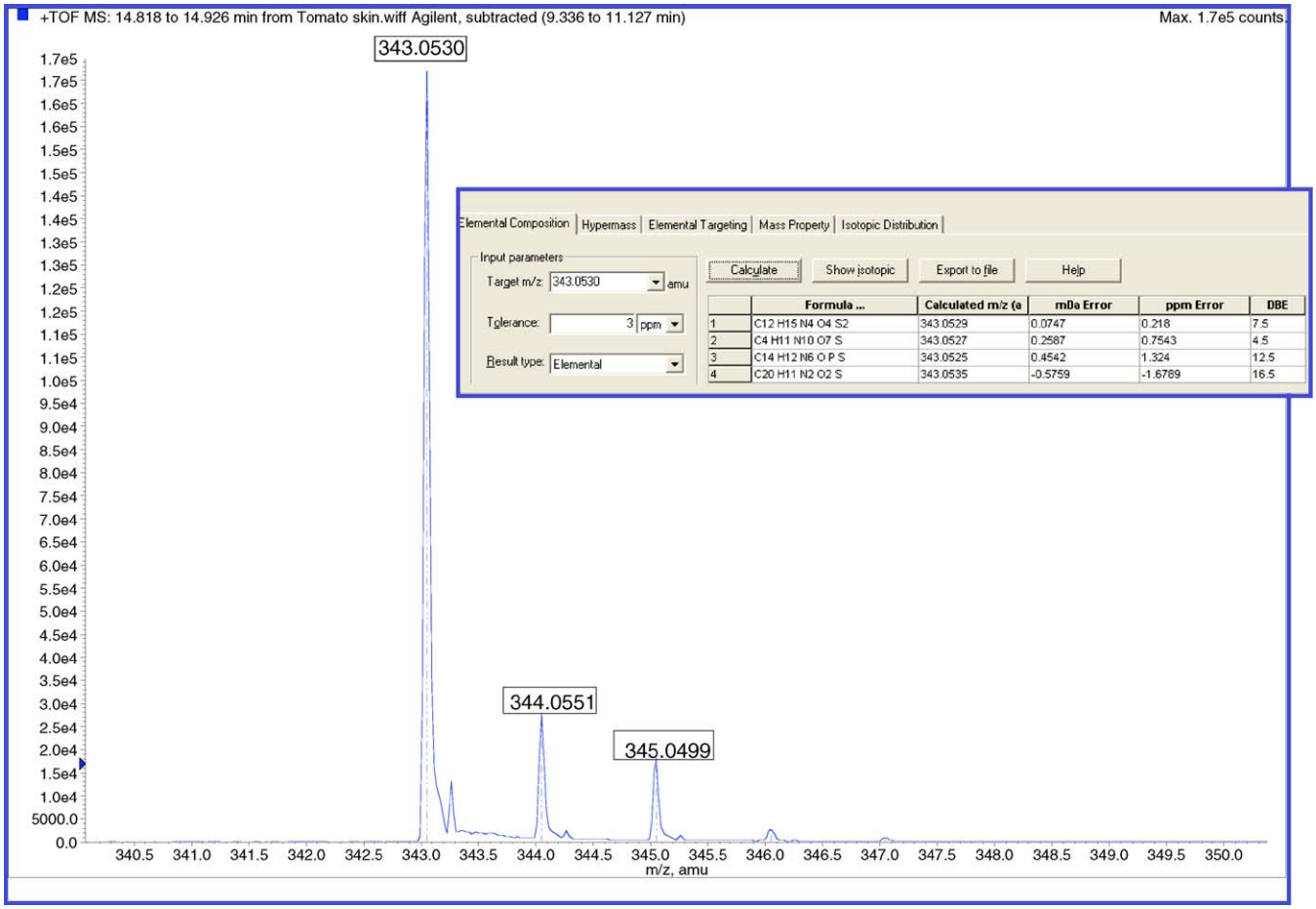

Fig. 2. LC/MS TOF Spectrum of peak 14.7 min.

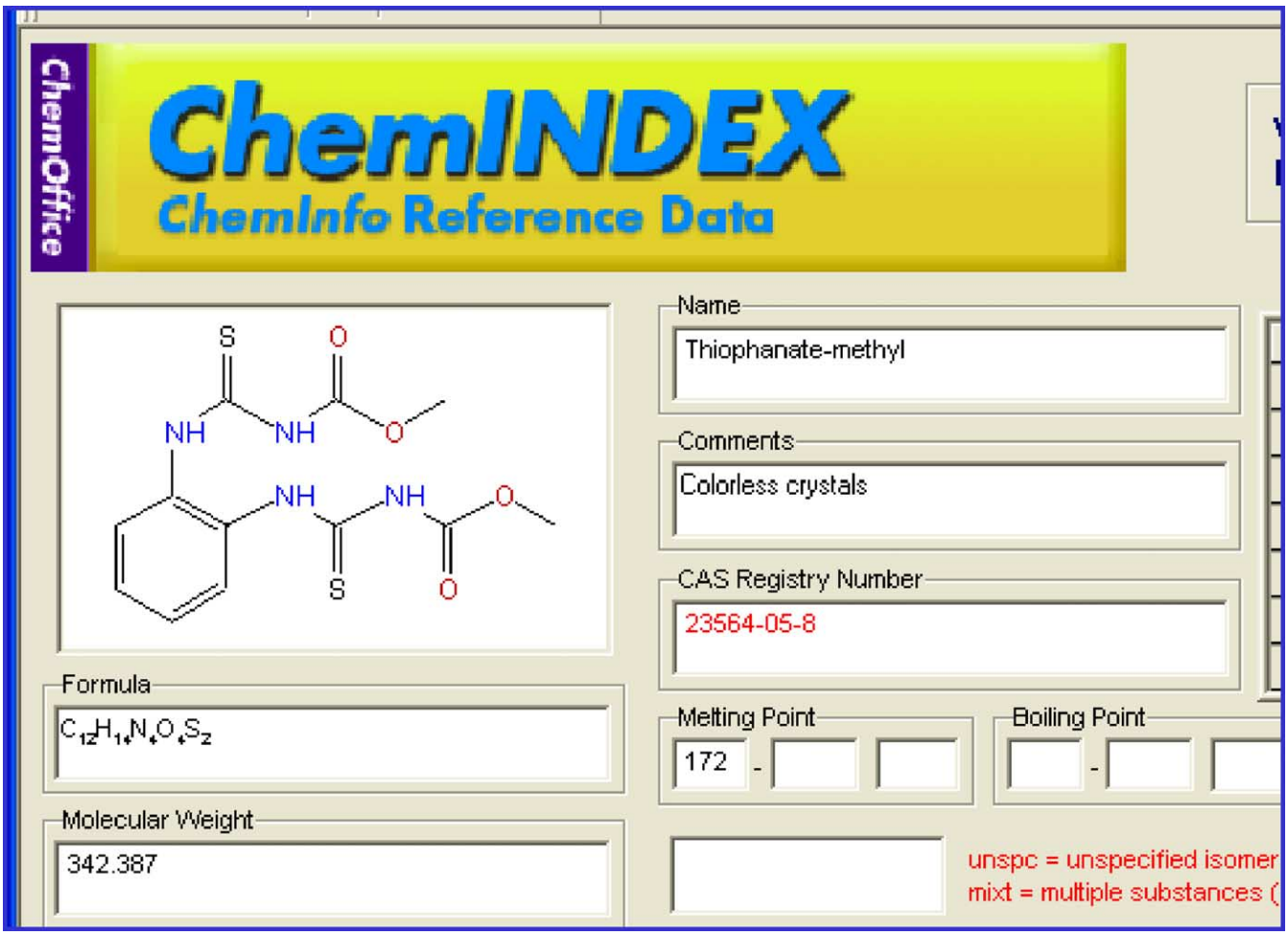

Fig. 3. Database search for empirical formula.

trap MS/MS spectrum of the *m/z* 343 ion. The spectrum contains two major peaks at *m/z* 311.0 and 150.8. Using a chemical drawing program, the two fragment ions may be easily drawn that result from a probable fragmentation based on the chemical structure of thiophanate methyl. These data give further important evidence that the database formula for the *m/z* 343.0530 is correct. Fig. 5 shows the CID spectrum of the accurate mass that was obtained from the LC/MS TOF. Note that there are ions at *m/z* 151.0321, 226.0644, 268.0211 (mass ion shown but not labeled), and 311.0267, all of which are consistent with the MS/MS spectrum show in Fig. 4. The two major ions have formula shown in Fig. 4 that match the fragment ions with an error of <0.1 ppm for the *m/z* 311 (the S-34 isotope for two sulfurs was also present, see Table 1) and −2.3 ppm for the *m/z* 150.8 ion (the S-34 isotope was present for one sulfur ion, see Fig. 5). These data are important because they give the confidence that the database analysis is correct for thiophanate methyl. Also the double bond and ring equivalents (DBE) for thiophanate methyl is 8, which is identical to the formula match in Fig. 2 (7.5 + 0.5 more for the lack of an electron in positive ion for a total of 8 DBE). The final step of the analysis was confirmation by authentic standard, which was carried out by accurate mass, MS/MS, and by chromatographic retention. All of which

Table 1
Measured mass, elemental composition, error and types of accurate-mass ions

Measured mass (*m/z*)	Elemental composition	Exact mass	Error (mDa)	Error (ppm)	Comments
192.0771	$C_9H_{10}N_3O_2$	192.0767	0.4	1.8	Carbendazim
160.0505	$C_8H_6N_3O$	160.0505	<0.0	−0.2	Carbendazim-methanol
306.1642	$C_{16}H_{24}N_3OS$	306.1634	0.8	2.4	Buprofezin
201.1059	$C_9H_{17}N_2OS$	201.1056	0.3	2.0	Buprofezin fragment ion
343.0530	$C_{12}H_{15}N_4O_4S_2$	343.0529	0.1	0.2	Thiophanate methyl
311.0267	$C_{11}H_{11}N_4O_3S_2$	311.0267	<0.1	0.1	Thiophanate methyl minus methanol
151.0321	$C_7H_7N_2S$	151.0324	0.3	−2.3	Basic thiophanate fragment ion

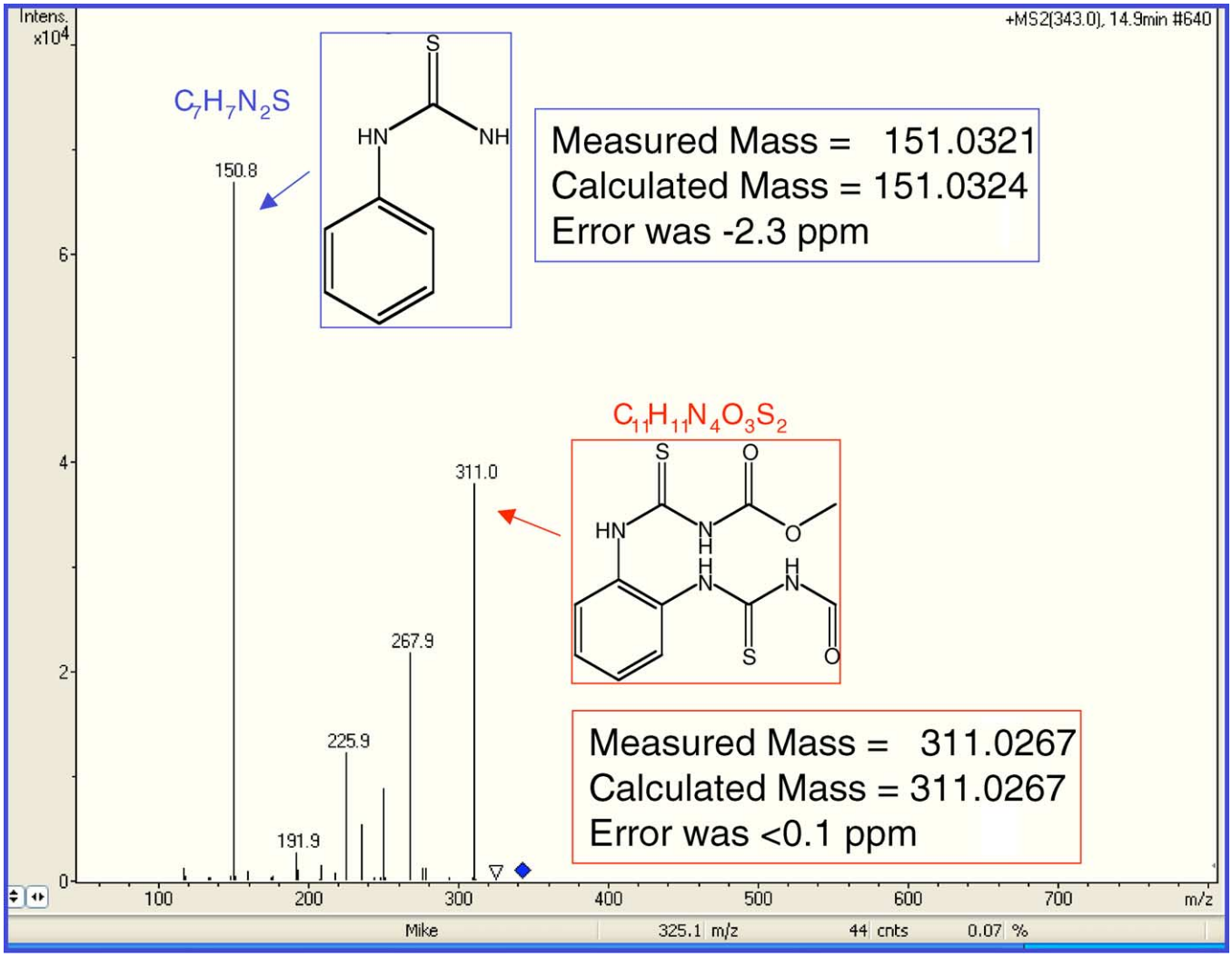

Fig. 4. LC/MS/MS ion trap spectra of the *m/z* 343 ion with accurate mass from LC/TOF/MS.

gave a positive identification for thiophanate methyl (data not shown).

Thus, it appears that this four-step procedure is a new approach and a powerful method for unknown analysis and is substantially different than the checking of library spectra that is commonly used in GC/MS identification methods or the use of selected ion monitoring or multiple reaction monitoring of LC/MS methods. The biggest liability of the approach is the lack of databases available to search empirical formulas. It is important to realize that accurate mass LC/MS techniques are ushering in a new approach to unknown identification especially when combined with LC/MS ion trap (or LC/MS Q/TOF) and chemical drawing software with accurate mass capabilities.

To check for its robustness (and repeatability) as a method, let us examine the peak at 3.2 min (Fig. 1). The accurate mass is 192.0771, which resulted in two empirical formulas from the calculator tool (no Cl or S was used because their isotopes at A + 2 were not present. The formula of $C_9H_{10}N_3O_2$ gave a database match in the Merck Index of carbendazim (Table 1) and the other formula gave no match. Carbendazim is a common fungicide that is used on fruits and vegetables and is known as a common degradation product of thiophanate methyl [14–15]. The ion trap MS/MS analysis of the *m/z* 192 resulted in a *m/z* 160 ion, which is the loss of 32 or methanol. This loss is consistent with the structure of carbendazim and with the accurate mass neutral loss from *m/z* 192.0071 to 160.0505 (Table 1), which is 32.0266 u, which is CH_3OH (accuracy of 0.6 mDa). Thus, one can look at either the accurate mass loss or at the fragment ion that is formed. Often there is only one match of the accurate mass loss because the mass is small and, therefore, there are many fewer matches for accurate mass formula. The final step of authentic standard gave a perfect match with carbendazim using LC/TOF/MS and LC/MS ion trap MS/MS (data not shown).

The last major peak in the chromatogram of Fig. 1 at 23.9 min resulted in the accurate mass of *m/z* 306.1642, which gave the empirical formula of $C_{16}H_{24}N_3OS$ at 2.4 ppm error. The empirical formula was then searched in the Merck Index and the insect growth regulator, buprofezin, was found.

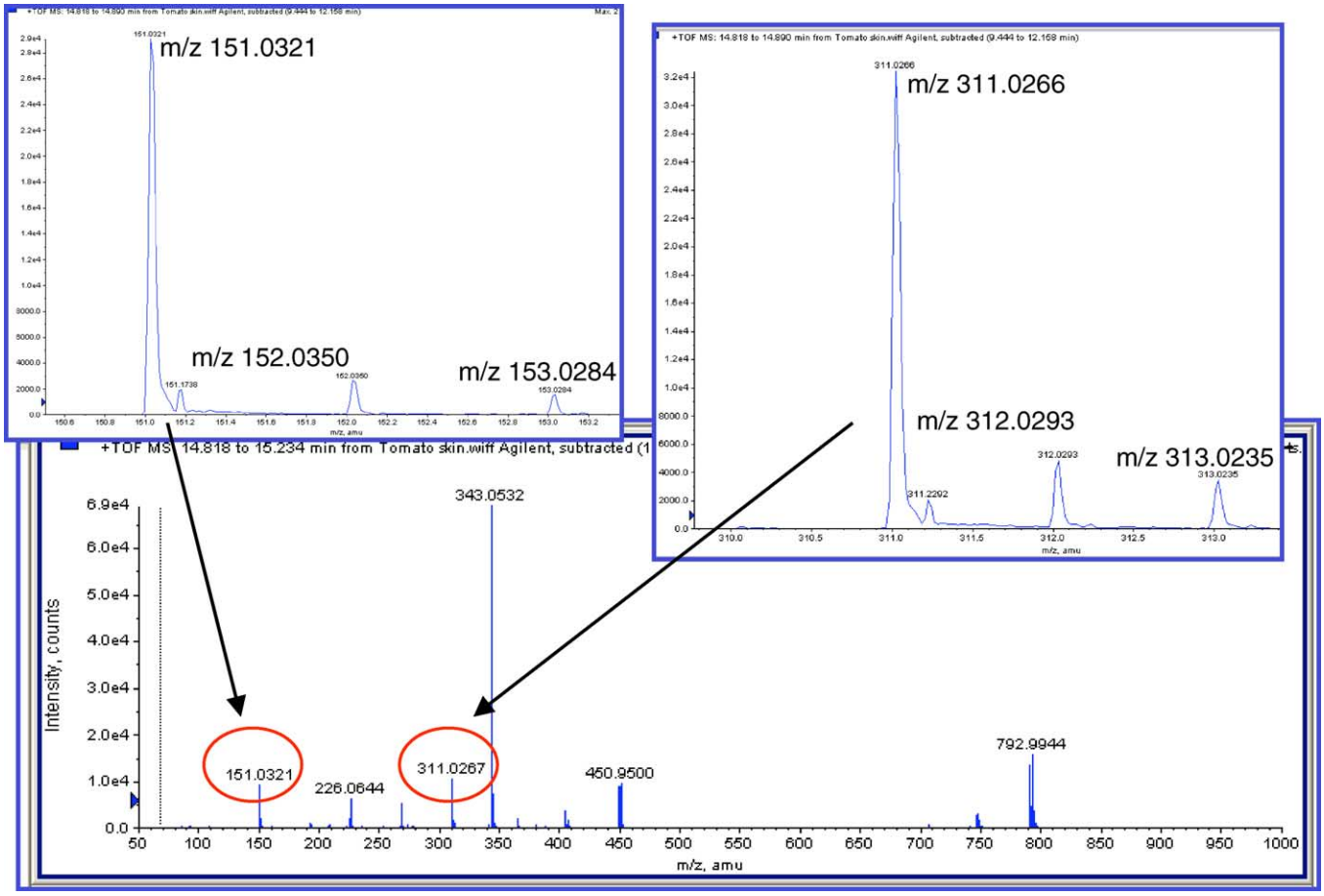

Fig. 5. Accurate mass empirical formula of the fragment ions using CID with sulfur isotopes shown.

Buprofezin is used extensively on white flies according to the Merck Index and according to a recent publication on tomatoes in Spain [16]. Thus, this compound was a good candidate for a positive identification by MS/MS. Note in Table 1 the ions at *m*/*z* 201.1059. LC/MSD ion trap MS/MS of the *m*/*z* 306 ion gave the 201 and further MS^3 gave the *m*/*z* 116 ion. It was possible to draw reasonable chemical structures for the 201 and 116 fragment ions that resulted from fragmentation of buprofezin. Furthermore, the accurate mass from the *m*/*z* 201.1059 fragment ion matched the formula from the chemical drawing software quite closely, which gave a higher certainty for identification. After obtaining the buprofezin standard, the final data show a perfect match, which further shows the ability of the LC/TOF/MS and LC/MS ion trap to identify unknowns. The amounts of these compounds in the tomato skin and the comparison of skin concentrations to whole tomato are the subject of another publication and will not be discussed here. Furthermore the quantitation ability of the LC/TOF/MS in complex matrices is also a subject for another publication.

Finally, in summary, the complimentary nature of the two instruments, LC/MS TOF and LC/MS ion trap, is shown. Other combinations of mass spectrometers that can be combined with TOF include Q/Trap and triple quadrupole for MS/MS verification, (and of course Q/TOF by itself). The limitations of these approaches were reviewed in an earlier publication [4], which includes a discussion of advantages of TOF and trap over Q/TOF alone (e.g. MS^n).

The method of identification described in this paper is limited in only one aspect, which is the size of the database that is being searched. At the moment, there are several databases that can be searched for pesticides. They include the Merck Index (∼10,000 compounds total, pesticides ∼500 compounds estimated), the ChemIndex (∼77,000 compounds total, pesticides ∼600 estimated) both from Cambridge Software, Cambridge, Massachusetts, and the websites of the Department of Agriculture of the USA [17] and pesticides of the UK [18]. While it is possible to identify compounds that are not in the database (new compounds, pesticide degradates, formulation impurities, and natural products), this is a more difficult task and not in the scope of this paper (addressed in reference [3] by the authors).

Finally, it has not escaped the authors attention that the well-known environmental analysis quote of Lynn Roberts [19] concerning the hunt for emerging contaminants (e.g. pesticides, pharmaceuticals in water, soil, food, etc.) by mass

spectrometry, "*As any analytical chemist knows, what you see depends on what you look for*" [19] is not ***always*** true (i.e. you do not always need standards and selected ions "a priori" to make identifications). The combined power of these two instruments (accurate mass within 3 ppm and MS/MS), chemical drawing software (including elemental calculators), and good databases make "a priori" unknown identification possible. Furthermore, it is important to take into account the mass of the electron in accurate mass calculations that are less than 2 ppm. Some of the instrument manufacturers have left this consideration out. Both the Agilent TOF and the Cambridge Software take the electron mass into account.

Finally, this procedure of unknown identification is also useful for other classes of compounds including pesticides in water and soil [3], pharmaceuticals in the environment [3–4], antibiotics in food and food products [11], and pesticide degradates in groundwater [8]. These are the subjects of both past and future publications for us and there is not space to thoroughly elaborate here on the many applications of this four-step approach.

Acknowledgments

We acknowledge assistance from the Agilent Staff, Dom Testa, Jerry Zweigenbaum, and Paul Zavitsanos of the US and Jaume Morales in Spain for the instrument set up and operation (LC/MS TOF). Mike Thurman acknowledges funding from the "Secretaria de Estado de Educacion y Universidades" of the Ministerio de Educacion y Ciencia. Imma Ferrer acknowledges her contract from the "Junta de Andalucia".

References

[1] EU Food Directives, 2002, 91/414/EEC.
[2] Food Quality Protection Act, 1998, (FQPA).
[3] Imma Ferrer, E.M. Thurman, Am. Chem. Soc. Symp. 850 (2003).
[4] Imma Ferrer, E.M. Thurman, Trends Anal. Chem. 22 (2003) 750.
[5] J.R. Gilbert, P. Lewer, D.O. Duebelbeis, A.W. Carr, C.E. Snipes, R.T. Williamson, Am. Chem. Soc. Symp. 850 (2003) 52, Chapter 4.
[6] S. Malato, L.P. Albanis, A. Aguera, D. Hernando, A. Fernandez-Alba, Am. Chem. Soc. Symp. 850 (2003) 66, Chapter 5.
[7] M.J. Benotti, P.L. Ferguson, R.A. Rieger, C.R. Iden, C.E. Heine, B.J. Brownawell, Am. Chem. Soc. Symp. 850 (2003) 109, Chapter 7.
[8] E.M. Thurman, I. Ferrer, E.T. Furlong, Am. Chem. Soc. Symp. 850 (2003) 109, Chapter 8.
[9] Y. Pico, C. Blasco, G. Font, Mass Spectro. Rev. 23 (2004) 45.
[10] I. Ferrer, C.E. Heine, E.M. Thurman, Anal. Chem. 76 (2004) 1437.
[11] E.M. Thurman, I. Ferrer, C.E. Heine, Anal. Chem. 76 (2004) 1228.
[12] A.A.K. Abou-Arab, Food Chem. 65 (1999) 509.
[13] W.J. Krol, T.L. Arsenault, H.M. Pylypiw Jr., M.J.I. Mattina, J. Agric. Food Chem. 48 (2000) 4666.
[14] M. Anastassiades, W. Schwack, J. Chromatogr. A 825 (1998) 45.
[15] A.R. Fernandez-Alba, A. Tejedor, A. Aguera, M. Contreras, J. Garrido, J. AOAC Int. 83 (2000) 748.
[16] A. Garrido-Frenich, F.J. Arrebola, M.J. Gonzaalez-Rodriguez, J.L.M. Vidal, N.M. Diez, Anal. Bioanal. Chem. 377 (2003) 1038.
[17] http://www.arsusda.gov/acsl/services/ppdb/.
[18] http://www.hclrss.demon.co.uk/index_mf_frame.html.
[19] L. Roberts, Am. Chem. Soc. Meet. (2000).