Canadian Society of Forensic Science Journal, 2014
Vol. 47, No. 2, 74–98, http://dx.doi.org/10.1080/00085030.2014.904605

Identification of unknown compounds from quadrupole GC-MS data using Cerno Bioscience MassWorks™

R.J. Abel*

Trace Evidence Section, Royal Canadian Mounted Police, NCFS-Alberta, 15707-118th Avenue, T5V 1B7, Edmonton, AB, Canada

Abstract

The MassWorks™ software package has been evaluated, in a forensic context, for use in conjunction with quadrupole GC-EI-MS data to identify unknown compounds. Methods were developed using this software to obtain the chemical formula for both molecular and fragment ions for three different dye molecules. These methods were then applied in a single-blind experiment to identify four unknown compounds. Mass accuracies well within 30 mDa were typical, and molecular formula candidate lists were developed without a reliance on complicated rule sets or manual formula assessments.

Keywords: GC-MS; quadrupole; accurate mass; MassWorks; forensic substance identification

Résumé

Le logiciel MassWorks™ a été évalué, dans un contexte de science judiciaire, pour l'utiliser avec des données d'un GC-EI-MS avec quadrupôle afin d'identifier des composés inconnus. Des méthodes ont été développées avec ce logiciel afin d'obtenir la formule chimique de molécules et de fragments d'ions de trois différentes teintures. Ces méthodes ont ensuite été utilisées lors d'une expérience à l'aveugle pour identifier quatre composés inconnus. Des masses précises à moins de 30 mDa des substances de référence étaient typiques, et des listes candidates de formules moléculaires ont été développées sans avoir recours à des règles compliquées ou à des évaluations manuelles de formules.

Mots-cléfs: GC-MS; quadrupôle; masse précise; Mass Works; identification de substance

Introduction

Many of the analytical challenges surrounding the identification of unknown substances are the same in the fields of pharmacology, toxicology, environmental chemistry and forensic science; however, because any material in existence may be submitted for forensic examination, the forensic chemist is likely to encounter a wider variety of substances than chemists working in other fields. In a forensic unknown identification, as in other fields where unknowns are encountered, a case history can help start the process of identifying or eliminating candidate compounds. Although caution must be exercised whenever excluding any compound from forensic consideration, similarity in the analytical approach to

*Email: robin.abel@rcmp-grc.gc.ca

unknown identification in every discipline makes new techniques in medical and environmental chemistry amenable to solving forensic problems and vice versa.

Gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) have both gained a place of prominence for the identification of unknowns since they first became readily available to researchers [1–3]. GC is the easiest separation technique to interface with mass spectrometry (MS), was the first widely available commercial separation and identification technique, and is currently the most used analytical instrument for the identification of small molecules [4, 5]. While the physical properties of large, charged, or thermally labile analytes generally demand other approaches, modern GC provides excellent resolution compared with other separation techniques. MS adds specificity to the selectivity of GC, and with the ever-increasing availability of compound databases both for purchase and available online, MS can be used for routine, accurate compound identification.

While high-resolution MS instrumentation (e.g. orbitrap and ion cyclotron resonance spectrometers) can provide sufficient information to determine the molecular formula for pure or well-separated compounds [6], the high cost of this instrumentation is the limiting factor in bringing high-resolution mass spectrometry into every lab. Despite the practical limitation of unit resolution, quadrupole mass analyzers are routinely used for laboratory analyses because of their low cost and robustness. In forensic applications, quadrupole instruments are successfully used for the identification of unknown compounds provided the candidates can be narrowed to a list of compounds with known or obtainable reference spectra, for example, a list of drug molecules and their known derivatives or metabolites in the context of an illicit drug identification [7]. When the list of possible candidate compounds is not easily narrowed (i.e., identification of a "true unknown"), compound identification can only be solved with higher resolution mass spectral data than is normally available from a quadrupole. In this study we demonstrate the use of the MassWorks^{TM} software package in conjunction with quadrupole GC-MS data to obtain the chemical formula for both molecular and fragment ions of "true unknowns" in a forensic context, providing a feasible alternative to more costly instrumentation.

While the quadrupole has a high mass error relative to other mass spectrometer designs [4, 8], even an instrument with high mass accuracy will yield a large number of candidate molecular formulae within a margin of error as low as ± 1 ppm [9]. Fortunately, more information than just the monoisotopic mass of an ion is available in the form of isotopic peaks which will vary in intensity and position based on the elements present in the compound. In the case of electron impact-mass spectrometry (EI-MS) the same mass and isotopic information available for the molecular ion will be available simultaneously for ions arising from fragmentation.

Modern quadrupole GC-MS software condenses the bulk of the instrumental data during acquisition through the use of peak centroids [8] because, until recently, even a modest data collection rate would overwhelm available computational power and data storage limits. The centroid process eliminates peak shape information, which leads to an increase in mass error during both instrument calibration and data acquisition; however, significant increases in both computational speed and storage now permit a more comprehensive approach to instrument calibration and data analysis. MassWorks uses the information present in raw, calibrated quadrupole data using an algorithm called Calibrated Lineshape Isotope

Profile Search (CLIPS) to generate a refined list of possible formulae for an ion, and uses the isotopic peak shape to assess the quality of fit for each proposed formula.

The algorithm used by MassWorks is based on the concept that mass spectrometric data, like chromatographic data, should be Gaussian with a variable peak width defined by the instrument resolution. Unlike spectroscopic techniques, the theoretical response of a mass spectrometer can be calculated exactly using the known isotopic abundances for the elements present in each calibrant ion, which allows the experimental response of the mass spectrometer to be compared directly to the theoretical response. By applying a suitable calibration correction, the response of the relatively low-resolution quadrupole mass spectrometer can be refined to give both greater mass accuracy and idealized peak shape which, in turn, allows comparison of simulated isotopic peak data to the experimental result for each analyte ion [8]. While MassWorks may be used with any mass spectrometer, including high resolution instruments [10], its ability to augment the standard single quadrupole may rapidly [11] provide invaluable information to the forensic scientist who is tasked with the identification of unknowns.

The work presented here demonstrates the potential for the use of the MassWorks software with single quadrupole GC-MS data for rapid and easy determination of the molecular formulae and identity of unknown compounds within a forensic context.

Materials and methods

All samples were analyzed using a Hewlett-Packard 6890 series gas chromatograph interfaced with an HP 5973 mass selective (quadrupole) detector using electron impact ionization (Agilent Technologies, Palo Alto, USA). The GC conditions were as follows: 30 m $\times$ 0.25 mm HP-1MS capillary column (0.25 μm film thickness) with hydrogen carrier gas held at a constant flow of 1.1 mL·min^{-1}; inlet temperature was set to 250°C with a split ratio of 20:1 at a split flow rate of 22.0 mL·min^{-1}. The GC column was initially held at 70°C for 2 min, ramped at 15°C·min^{-1} to 320°C and held for 6 min. After the hold, the column was cooled to 70°C, held for 4 min, and the perfluorotributylamine (PFTBA) valve opened during the final minute to introduce an internal calibration peak into each chromatogram. The mass spectrometer was set to an ionization potential of 70 eV, and to acquire raw scan data at a threshold of zero counts between 40 and 400 m/z with eight measurements taken at each mass (scan rate of approximately 2 per second).

Post-processing of GC-MS data was performed using the MassWorks software package version 2.0 (Cerno Bioscience, Danbury, CT, USA). MassWorks calibration was performed using PFTBA fragment ions of reasonable intensity and without spectral interference from overlapping ions (see Table 1). The molecular formulae of the PFTBA fragments were entered into the MassWorks calibration table with the corresponding m/z regions marked automatically by the software, followed by manual refinement to encompass only the region occupied by the ion peaks. Mass spectra for analyte peaks were generated in MassWorks using spectral averaging for the entire peak in the absence of co-elution, or for a section of the peak that was free from interferences. Background subtraction was also applied to any peak where sufficient adjacent baseline could be averaged to generate a suitable background spectrum. Where analyte peak intensity was more

Table 1. PFTBA fragments used for calibration of MassWorks.

Formula	Mono isotope (m/z)	Closest centroid	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE
CF3	68.9952	68.9963	1.1486	16.6	99.8	23.2
C2F4	99.9936	99.9937	0.1081	1.1	99.6	11.3
C2F5	118.9920	118.9929	0.8918	7.5	99.6	8.8
C3F5	130.9920	130.9940	1.9602	15.0	99.7	25.7
C3F6	149.9904	149.9917	1.3231	8.8	99.6	1.7
C3F7	168.9888	168.9910	2.1610	12.8	99.8	1.3
C4F6N	175.9935	175.9943	0.8438	4.8	99.3	2.6
C4F7	180.9888	180.9930	4.1821	23.1	99.6	1.3
C4F8N	213.9903	213.9926	2.3113	10.8	96.2	16.5
C4F9	218.9856	218.9945	8.9262	40.8	99.1	72.1
C5F9	230.9856	230.9863	0.6855	3.0	99.2	0.8
C5F10N	263.9871	263.9917	4.6031	17.4	99.6	11.1
C6F12N	313.9839	313.9880	4.1000	13.1	99.0	0.9
C7F12N	325.9839	325.9885	4.5504	14.0	98.6	0.8
C8F14N	375.9807	375.9850	4.3000	11.4	99.1	0.6

than double that of the PFTBA calibrant, only the leading or tailing edge of the peak was averaged to avoid ion intensity compression, yielding a better theoretical match to the experimental peak. Calibration was deemed successful based on the criteria laid out by the software developer [12].

The minimum and maximum values for each element in the CLIPS search were defined using limits determined from searches of the NIST Chemistry WebBook [13] online compound database. The remaining default CLIPS parameters were optimized by assessing the position of the correct formula for known dye compounds in the search results. The optimized settings are listed in Table 2.

2-(methylamino)anthraquinone (MAAQ) was obtained from Sigma Aldrich (St. Louis, Missouri, USA), Indigo dye was obtained as a natural product (Maiwa Dyes, Vancouver, British Columbia, Canada), and hexamethyl-p-rosaniline chloride ('Gentian Violet') was obtained from London Drugs (Edmonton, Alberta, Canada). All three were dissolved together in dichloromethane (Fisher Scientific, Hampton, New Hampshire, USA) at a concentration of approximately 1 $mg \cdot mL^{-1}$ each, and subjected to GC-MS without further preparation.

Table 2. Optimized CLIPS search parameters.

Parameters	Value	Comment
Mass tolerance (mDa)	30	Increase if necessary
Electron state	Both	Set to odd for the molecular ion if highly confident
DBE range min	−0.5	
DBE range max	30	Maximum unsaturation of a C_{30} hydrocarbon
Profile mass start (Da)	−1.5	
Profile mass end (Da)	3.5	
Interference rejection	0.0035	
Mixture search	Empty	Populate with formulae of interfering ions

Four unknown compounds (n,n-diethyl metatoluamide (DEET); pseudoephedrine; ethyl 4-aminobenzoate; and strychnine) were supplied by another scientist, single-blinded, and were obtained from the RCMP chemical reference collection. All four compounds are available for purchase from both Sigma Aldrich (St. Louis, Missouri, USA) and Fisher Scientific (Hampton, New Hampshire, USA).

The NIST Chemistry WebBook was also consulted for the average retention index of each compound on HP-1MS equivalent capillary columns, and compared with experimental values calculated using the Van den Dool/Kratz method [14–16] as a final test of each identification.

Results and discussion

Validation of MassWorks using selected dyes

Three forensically interesting dyes were selected as heterocyclic standards with which to validate the function of the MassWorks software (Figure 1). MAAQ is frequently encountered in casework involving bank dye packs and is routinely analysed by GC-MS [17]. Indigo, (2,2'-biindoline)-3,3'-dione, is a popular dye used in dark coloured fabrics, which chromatographs poorly on the DB-1 equivalent column used. Gentian Violet, hexamethyl-p-rosaniline chloride, is a relatively large molecule, late eluting, exists as an ion in solution, converting to the leucobase form to attain the gas phase, and is used extensively as a taggant, a fabric dye, a pigment in dark inks, and as an antimicrobial/anthelmintic agent.

Instrument and software setup

The three dyes were separated by GC, with elution of MAAQ at 14.3 min, Indigo surrounding 16.8 min, and Leucogentian Violet at 18.8 min. The peak shape of Indigo was noticeably distorted, most likely due to the use of a DB-1 equivalent column. A separation of the dyes was attempted on a DB-WAX equivalent column; however, the significantly lower thermal tolerance of the column stationary phase required lower column temperatures, which prevented any of the dyes from eluting.

While a single quadrupole can generate a spectrum from compounds with a mass up to ~4000 m/z [4], GC is generally limited to heat-stable compounds that

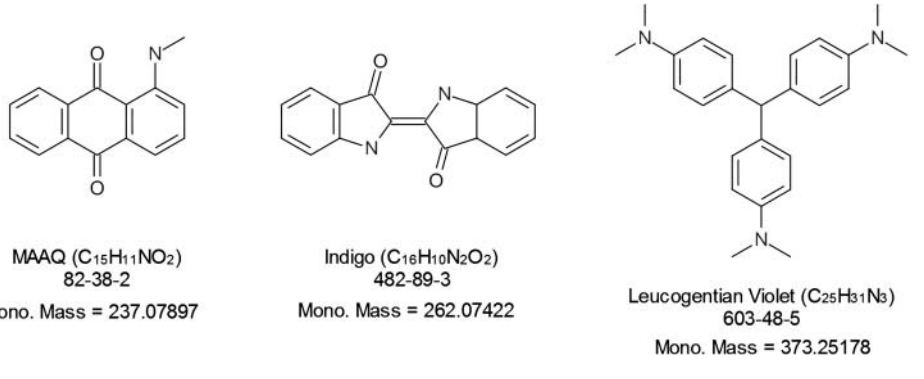

Figure 1. Dye structures, selected for validation of MassWorks.

do not exceed a 500°C boiling point and/or a molecular weight of 1000 g·mol^{-1} [18]. Accordingly, thermally labile compounds, and those that are ionic, high in mass or low in volatility are not suitable for separation by GC, and are generally subjected to a technique other than GC-MS (e.g. LC-MS, FT-IR, Raman, Py-GC-MS, XRD, etc.). Furthermore, given that our GC inlet temperature is set to 250°C, and the column temperature does not exceed 320°C, compounds with a boiling point above approximately 300°C may be excluded from consideration. In addition, in order to limit the final size of the data file, the m/z range of the MS is set to a maximum of 400 m/z in our work, effectively excluding compounds with a mass of greater than 400 g·mol^{-1}. While compounds with a molecular ion above 400 m/z will certainly be encountered in unknown identifications, it is anticipated that the investigator would identify this situation either during a screening injection using a wide m/z scan range or during interpretation of the spectrum, and would re-inject the sample with an appropriately increased m/z scan range, at which point the CLIPS search parameters would be expanded accordingly.

While PubChem [19] and ChemSpider [20] have been used by others for unknown identification [21, 22], neither were found to be suitable for assisting in the identification of a general unknown. PubChem is excellent for searching a candidate list of molecular formulae, but does not allow open-ended compound searches based solely on physical properties, nor does it have the capability to define a molecular formula using wildcards or specified values for individual elements. At approximately 26 million records, ChemSpider is one of the largest compound databases available; however, open-ended searches based on physical properties required excessive search times (hours for a single search), depended largely on calculated properties rather than experimental ones, and yielded such extensive hit lists that it was not possible to retrieve the hits as a list of formulae that could be used to define reasonable search criteria for CLIPS. Furthermore, the ChemSpider web interface generated errors when searching by the molecular weight property field in conjunction with the calculated boiling point property field, and would only permit searching molecular formula by exact match rather than by specified element values or by wildcards. While PubChem and ChemSpider are expected to be excellent resources for evaluating the formula lists provided by CLIPS, the NIST Chemistry WebBook [13] was relied upon for defining the CLIPS search criteria in spite of its much smaller compound database (~70,000 records based on last posted count as of June 2005).

While the NIST Chemistry WebBook is not a comprehensive list of known compounds, it is reasonable to assume that the database provides an average cross-section, such that the molecular formulae are representative of the limits one can expect to encounter for an unknown. The NIST Chemistry WebBook parameters for molecular weight searches permitted the use of wildcards in the molecular formula field, although the search could not be parameterized based on boiling point. The NIST Chemistry WebBook species list was filtered to exclude all compounds containing any element other than C, H, N, O, F, Si, P, S, Cl and Br, followed by exclusion of all compounds not within the mass range of 40 to 400 g/mol. The resulting filtered compound list was searched for the number of compounds containing 2-3, 4-5, 6-7, 8-9, 10-15, 16-20, 21-30, 41-50, 51-75 and 76-100 of each element of C, H, N, O, F, Si, P, S, Cl and Br (see Table 3a for results), where the total number of compounds returned for each search was used

Table 3a. Number of compounds containing only these elements in the NIST Chemistry WebBook (between 40 and 400 g/mol).

	2–3	4–5	6–7	8–9	10–15	16–20	21–30	31–40	41–50	51–75	76–100	Max #
H	1402	2443	3527	4441	15741	9900	7565	2459	793	168	0	58
C	1880	3858	7240	8447	18882	6123	2375	2	0	0	0	32
N	7022	1206	230	20	18	0	0	0	0	0	0	14
O	16198	3872	738	173	93	6	0	0	0	0	0	17
F	1513	473	297	102	83	1	0	0	0	0	0	16
Si	801	37	4	2	4	0	0	0	0	0	0	13
P	92	28	3	0	0	0	0	0	0	0	0	7
S	1290	138	8	3	2	0	0	0	0	0	0	11
Cl	2315	456	146	8	0	0	0	0	0	0	0	8
Br	454	11	0	0	0	0	0	0	0	0	0	4
I	55	0	0	0	0	0	0	0	0	0	0	3

Table 3b. Optimized initial elemental maxima for CLIPS searches.

Element	Max	Comment
C	30	
H	62	
N	4	
O	8	
Si	4	Exclude if numerous unrealistic formulae
S	2	
Cl	4	
Br	2	
P	1	Excluded unless other data suggests presence
F	8	Excluded unless other data suggests presence

as an indication of the likelihood of encountering an unknown containing an amount of each element in that range. By following this approach, a scientist may decide what elemental maxima to use in the CLIPS search parameters to avoid unrealistic formulae without excluding any compounds of possible forensic importance. Metal complexes and ions could not be excluded from the NIST Chemistry WebBook when searching by molecular weight, resulting in artificially inflated compound lists; however, this was considered to be of benefit in overestimating the likelihood of each given elemental maximum, and to provide the most inclusive search criteria. The elemental maxima specified for CLIPS searches in our work are listed in Table 3b.

In normal casework, the scientist would be expected to have some idea of the type of compound present based on case history and the results of a simple library search for the unknown. This information would be used prior to employing MassWorks to further restrict the element limits. Ultimately, the elemental search criteria are set at the discretion of the scientist, and care must always be exercised to avoid excluding valid formulae because exclusion of the correct molecular formula will result in an inability to assign the fragmentation pattern. When evaluating the results of numerous CLIPS searches, the inclusion of F and P resulted in large numbers of CLIPS hits with unrealistic formulae, and so the decision was made to omit these two elements from the CLIPS element list unless their presence was suspected or suggested by other experimental evidence.

Double bond equivalents (DBE) were restricted to a minimum of –0.5 (according to McLafferty's guidelines [6]), and a maximum of 30, which is the DBE for a straight chain C_{30} hydrocarbon with 15 triple bonds. The remaining CLIPS settings (mass tolerance, profile mass range, electron state, and interference rejection) may be determined iteratively by the user by searching the molecular ion for a known compound, where improvements in spectral accuracy and mass error for the correct formula relative to other candidate formulae indicates an improvement in settings. In this work, mass accuracy appeared to decrease with increasing retention time, possibly due to interference from column bleed at higher temperatures. As a result, a relatively high mass tolerance of 30 mDa was chosen to avoid excluding correct formulae; however, this was still a significant improvement over the mass tolerance of a traditional quadrupole and should not significantly complicate an unknown identification due to the discrimination provided by

spectral fitting. Specifying the CLIPS mass tolerance in mDa was preferred over ppm because CLIPS searches on fragment ions resulted in larger mass errors for lighter fragments, which often resulted in the exclusion of the correct formula even when set as high as 100 ppm.

Significant restriction of CLIPS candidate formulae was achieved by setting the electron state for an ion either to ODD or EVEN, and in cases where the scientist is quite sure of the state of an ion, it may be used to great advantage. For example, the molecular ion is always ODD in EI-MS if it is sufficiently stable and has been identified correctly [6]. However, unless one is very confident of the electron state of an ion, the parameter is best left set to BOTH to ensure inclusion of all valid formulae. Profile mass range was always set from –1.50 to 3.50, affording a better visual assessment of spectral accuracy, especially when the mixture search fields were used. Interference rejection is applied by the software when accounting for the contents of the mixture search fields in the theoretical fit, and was found to operate best between the values 0.0015 and 0.0035; however, this result is purely empirical and may not hold true for all analytes or instrument conditions.

The background subtraction feature is important for increasing spectral accuracy when column bleed becomes significant or coelution is present. Coelution can be difficult to determine when a peak is fully concealed beneath a larger one, or when peak shapes are not Gaussian due to poor chromatography. Tools such as Chemstation peak purity, or the free NIST software AMDIS [23] can be used to rapidly profile an entire chromatogram to define coelution or spectral inhomogeneity, although their applicability and use is beyond the scope of this paper. As a general rule, background subtraction should be used where a stable baseline, free from interference, is present adjacent to the peak of interest. When using background subtraction, special attention following subtraction must be paid to the mass spectrum of both the analyte and the surrounding baseline. The presence of a fluctuating baseline or negative peaks in either spectrum may indicate that the background being subtracted has not been well selected. Since it is critical to spectral accuracy that legitimate intensity not be subtracted from the ion being profiled, in situations where any doubt regarding background subtraction exists, it is better to omit the subtraction altogether and instead to attempt to account for ion interferences using the CLIPS mixture search.

In this work, we found that the mixture search fields are very powerful when used appropriately. In a case where the molecular ion becomes more stable after the loss of a hydrogen, the actual $M^{\bullet+}$ may be obscured by the more intense $[M\text{-}H]^{\bullet+}$ peak, in which case the spectral fit proposed for the top hit will fit the isotopic peaks poorly. In this case, the overlap must be accounted for, and the inclusion of '-H' as the mixture formula will result in appropriate CLIPS results and a far better spectral fit. These fields can also be advantageous when the molecular formulae for coelution or column bleed are included in the mixture search for peaks of interest.

Application of MassWorks to MAAQ, Indigo, and Gentian Violet

Retention times for each dye were first identified by acquiring the sample under routine conditions (i.e. normal scan) followed by peak identifications by

searching each spectrum against the NIST 02 database. The sample was then re-analysed with the MS in raw scan mode to provide the data necessary for MassWorks. Upon opening the raw scan data in MassWorks, a new calibration was generated using PFTBA fragments of suitable intensity. Any PFTBA fragments yielding a spectral match of 99.0% or greater were retained, while any fragments falling below this threshold were excluded and the calibration repeated. The final list of fragments used for calibration is provided in Table 1. The calibration was then applied to the chromatographic data, and each of the three dye peaks marked by a retention time window. Before evaluating each windowed dye peak, the baseline immediately prior to the peak was examined and, if suitably absent of eluting material, a second window of approximately the same width was marked and used for background subtraction.

MAAQ provided a well-defined molecular ion at 237.0763 m/z from which the correct molecular formula was easily obtained using the optimized CLIPS settings (Table 2). The presence of a pronounced hump, one m/z unit below the monoisotopic peak, suggested the loss of a single hydrogen from the molecular ion. CLIPS was run again, including the relative formula '-H' in the mixture search field to account for the contribution of both ions to the monoisotopic peak, resulting in an excellent spectral fit and the correct formula ranked as the top candidate by spectral accuracy (Figure 2, Table 4). The ion type in CLIPS was set to include all even and odd electron formulae, in spite of a high degree of confidence that the molecular ion had been correctly identified. While using the ODD ion setting would have provided the shortest list of possible formulae, it is more prudent to use the setting BOTH to ensure that a well-defined even-electron fragment ion is not mistaken for the molecular ion in cases where the true molecular ion may be too weak to be visible or is altogether absent.

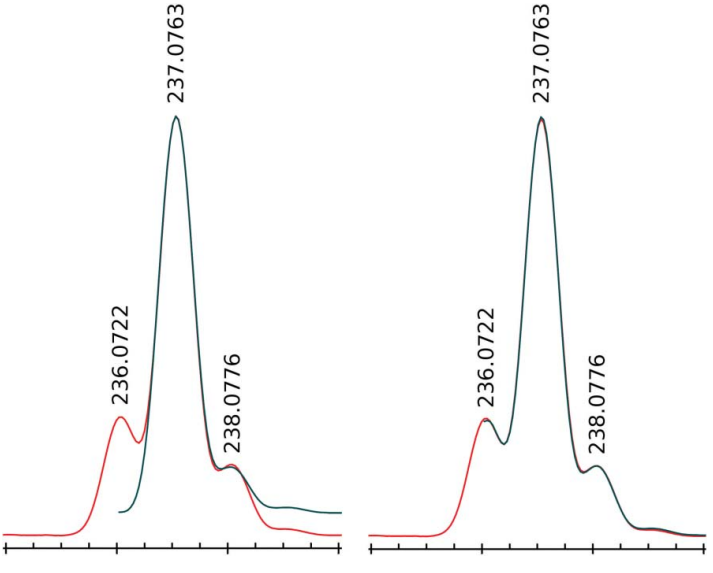

Figure 2. CLIPS spectral fits for MAAQ ($C_{15}H_{11}NO_2$) with no mixture formula specified (left) and a mixture formula of –H (right).

Table 4. MAAQ (-H mixture search).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C15H11NO2	*237.0790*	*2.6786*	*11.2985*	*99.1132*	*42*	*11.0*
C14H11N3O	237.0902	13.9120	58.6815	99.1047	43	11.0
C14H9N2O2	237.0664	−9.8975	−41.7480	98.8938	53	11.5
C13H9N4O	237.0776	1.3359	5.6350	98.8028	57	11.5
C19H9	237.0704	−5.8747	−24.7798	97.5444	117	15.5
C9H13N4O2Si	237.0808	4.4772	18.8851	97.3924	113	6.5
C10H13N2O3Si	237.0695	−6.7562	−28.4979	97.3165	116	6.5
C11H15NO3Si	237.0821	5.8199	24.5486	97.3065	116	6.0
C12H13O5	237.0763	−0.0015	−0.0062	97.1154	137	6.5
C11H13N2O4	237.0875	11.2319	47.3768	96.9838	144	6.5

Note: Italics show the correct formula for the target molecule.

Indigo chromatographed poorly, exhibiting an irregular peak shape. Nevertheless, integrating the entire elution region provided a spectrum with a molecular ion at 262.0623 m/z for which CLIPS yielded the correct molecular formula. No significant fragment ion interference was present, but the relative formula '-H' was still included in the search parameters to account for the small peak one m/z unit below the monoisotopic peak. The ion setting for CLIPS was once again set to BOTH, resulting in a good spectral fit and the correct formula ranked as the top candidate by spectral accuracy (Figure 3, Table 5). It is noteworthy that even in cases where an analyte does not chromatograph well, it may still be possible to identify the molecular formula with MassWorks provided it can be injected and eluted in some fashion.

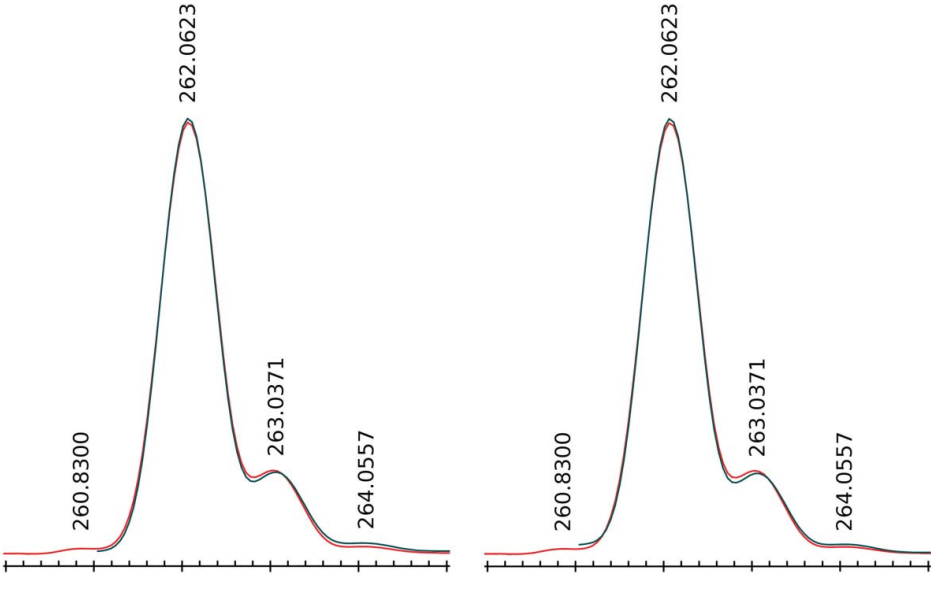

Figure 3. CLIPS spectral fits for Indigo ($C_{16}H_{10}N_2O_2$) with no mixture formula specified (left) and a mixture formula of –H (right).

Table 5. Indigo (-H mixture search).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C16H10N2O2	*262.0742*	*11.9276*	*45.5143*	*98.1273*	*34*	*13.0*
C17H10O3	262.0630	0.6942	2.6489	98.1003	35	13.0
C16H8NO3	262.0504	−11.8819	−45.3399	97.9542	37	13.5
C15H8N3O2	262.0617	−0.6485	−2.4745	97.9355	38	13.5
C14H6N4O2	262.0491	−13.2245	−50.4634	97.6463	43	14.0
C19H6N2	262.0531	−9.2018	−35.1130	97.4931	46	18.0
C20H8N	262.0657	3.3743	12.8758	97.0930	53	17.5
C14H10N4Si	262.0675	5.1729	19.7391	96.3638	60	13.0
C13H14O4Si	262.0661	3.8355	14.6357	96.2815	62	8.0
C11H12N3O3Si	262.0648	2.4928	9.5122	96.2443	62	8.5

Note: Italics show the correct formula for the target molecule.

The final dye, *Leucogentian Violet*, provided a well-defined molecular ion at 373.2361 m/z with a pronounced hump at one m/z unit below the monoisotopic peak; however, even after including the relative formula '-H' in the CLIPS mixture field, the default 10 mDa mass tolerance excluded the correct formula from the candidate list, as did an increased value of 15 mDa (Table 6a). A value of 20 mDa was found to be sufficient to include the correct formula within the top three candidates with a very good spectral fit (Table 6b, Figure 4). After increasing the mass tolerance, and including the relative formula '-H' in the mixture field, the correct formula was included in the candidate list, but was still not presented as the best match. However, because the molecular ion was well defined, the ODD ion setting was used and gave the correct formula as the best hit (Table 6c).

Based on the results from the analysis of Leucogentian Violet, it was decided that 30 mDa should be used in the CLIPS settings to provide a margin of safety for the identification of unknown compounds. It appears that later eluting peaks may be subject to greater mass error due, in part, to contributions from greater column bleed towards the end of a run. If a reasonable formula identification

Table 6a. Leucogentian Violet (-H mixture search, 15 mDa tolerance).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C24H29N4	373.2392	3.1220	8.3645	98.7809	114	12.5
C25H29N2O	373.2280	−8.1114	−21.7327	98.7018	116	12.5
C26H31NO	373.2406	4.4646	11.9619	98.5472	129	12.0
C23H33O4	373.2379	1.7845	4.7813	97.9621	179	7.5
C22H33N2O3	373.2491	13.0179	34.8786	97.8278	193	7.5
C22H31NO4	373.2253	−10.7915	−28.9134	97.6151	211	8.0
C21H31N3O3	373.2365	0.4419	1.1839	97.4663	227	8.0
C20H33N4OSi	373.2424	6.2632	16.7809	97.2783	232	7.5
C20H29N4O3	373.2240	−12.1342	−32.5107	97.2294	260	8.5
C21H33N2O2Si	373.2311	−4.9702	−13.3164	97.0589	251	7.5

Table 6b. Leucogentian Violet (-H mixture search, 20 mDa tolerance).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C24H29N4	373.2392	3.1220	8.3645	98.7809	114	12.5
C25H29N2O	373.2280	−8.1114	−21.7327	98.7018	116	12.5
C25H31N3	*373.2518*	*15.6980*	*42.0592*	*98.6860*	*120*	*12.0*
C26H29O2	373.2168	−19.3448	−51.8300	98.5883	124	12.5
C26H31NO	373.2406	4.4646	11.9619	98.5472	129	12.0
C27H33O	373.2531	17.0407	45.6566	98.2825	151	11.5
C23H33O4	373.2379	1.7845	4.7813	97.9621	179	7.5
C22H33N2O3	373.2491	13.0179	34.8786	97.8278	193	7.5
C22H31NO4	373.2253	−10.7915	−28.9134	97.6151	211	8.0
C21H31N3O3	373.2365	0.4419	1.1839	97.4663	227	8.0

Note: Italics show the correct formula for the target molecule.

cannot be reached with the initial CLIPS settings, expanding the mass tolerance is recommended as the best first step in identifying the correct formula. While any increase in mass tolerance results in the inclusion of a substantial number of additional possible formulae [9], this will not significantly increase the difficulty of the identification because the discrimination of candidate formulae by the CLIPS spectral matching procedure ensures that only formulae with a good experimental fit are added to the top results without requiring the user to engage in time-consuming manual evaluation of each additional candidate [24].

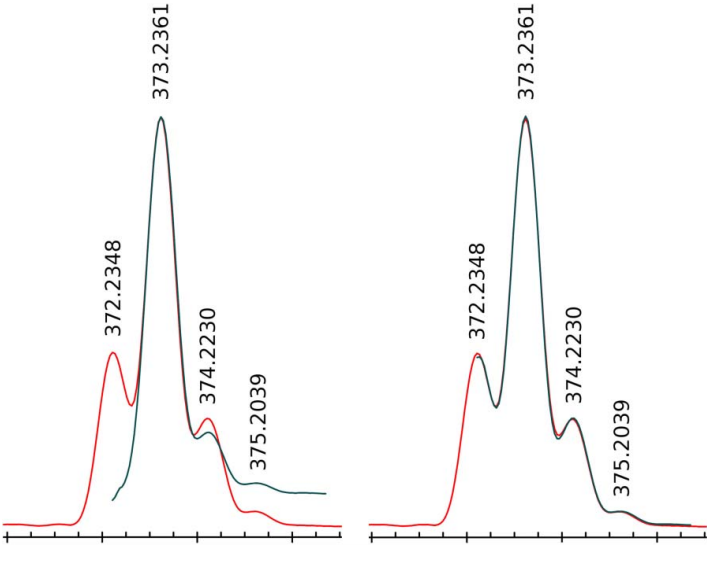

Figure 4. CLIPS spectral fits for Leucogentian Violet ($C_{25}H_{31}N_3$) with no mixture formula specified (left) and a mixture formula of –H (right).

Table 6c. Leucogentian Violet (-H mixture search, 20 mDa tolerance, odd ions only).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C25H31N3	*373.2518*	*15.6980*	*42.0592*	*98.6860*	*120*	*12.0*
C26H31NO	373.2406	4.4646	11.9619	98.5472	129	12.0
C22H31NO4	373.2253	−10.7915	−28.9134	97.6151	211	8.0
C21H31N3O3	373.2365	0.4419	1.1839	97.4663	227	8.0
C20H31N3O2Si	373.2186	−17.5462	−47.0110	97.2514	235	8.0
C21H35N3OSi	373.2549	18.8393	50.4755	97.0518	252	7.0
C17H35N3O4Si	373.2397	3.5832	9.6003	96.9675	260	3.0
C18H35NO5Si	373.2285	−7.6502	−20.4970	96.9518	260	3.0
C22H35N3S	373.2552	19.0688	51.0905	96.8897	265	7.0
C23H35NOS	373.2439	7.8354	20.9933	96.8063	273	7.0

Note: Italics show the correct formula for the target molecule.

Identification of unknowns: results of blind tests

For the purposes of truly challenging the MassWorks software, and ensuring that the author remained appropriately unbiased, no MS library searches were carried out on the unknown compounds. After validating and optimizing CLIPS for the known dyes, four unknowns were evaluated in a single-blind experiment.

Blind test 1: single unknown compound 'A42'

The first unknown was provided as a clear, colourless methanol solution and chromatographed as a single peak eluting at 9.0 min, with the heaviest ion at 191.1337 m/z. The intensity of the 191 m/z peak relative to the 190 m/z peak was too large to be the result of an [M+1] isotopic peak alone, suggesting that the 190.1202 m/z ion was generated by a loss of one hydrogen from 191 m/z. While the NIST 02 mass spectral database was available, to ensure an unbiased evaluation of the MassWorks software no search of the unknown was performed. A preliminary CLIPS search on the 191 m/z peak provided very poor spectral fits for the top 10 candidates (Table 7a). Including the relative formula '-H' in the mixture field resulted in far better spectral fits (Figure 5), with improved spectral

Table 7a. Unknown A42 (no mixture search).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C10H23OS	191.1470	13.2610	69.3810	48.0996	382	−0.5
C9H23O2Si	191.1467	13.0315	68.1801	47.8861	389	−0.5
C7H17N3O3	191.1270	−6.7086	−35.0988	26.2651	653	1.0
C10H15N4	191.1297	−4.0285	−21.0769	26.2447	663	5.5
C8H19N2O3	191.1396	5.8675	30.6983	26.1956	654	0.5
C11H17N3	191.1422	8.5476	44.7203	26.0366	660	5.0
C9H19O4	191.1283	−5.3659	−28.0741	26.0216	650	0.5
C12H17NO	*191.1310*	*−2.6858*	*−14.0521*	*25.8553*	*657*	*5.0*
C13H19O	191.1436	9.8902	51.7451	25.7817	657	4.5
C6H19N4OSi	191.1328	−0.8872	−4.6419	24.9206	632	0.5

Note: Italics show the correct formula for the target molecule.

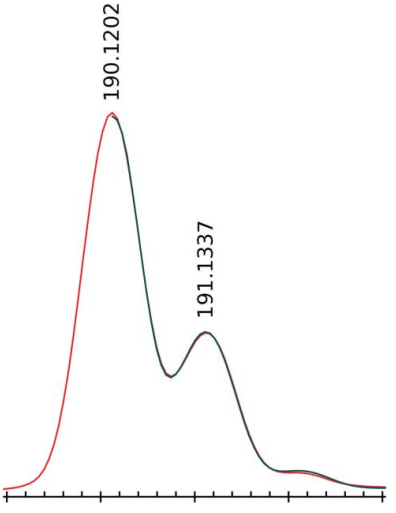

Figure 5. CLIPS spectral fit for Unknown A42 for $C_{12}H_{17}NO$ (the smaller peak at 191 m/z) with a mixture formula of –H.

accuracy scores for the 10 formula candidates (Table 7b). The addition of the search restriction to ODD ions resulted in the refined list of candidates presented in Table 7c. Formulae having silicon along with heteroatoms other than oxygen were deemed unlikely, as were those with many heteroatoms but low or no double-bond equivalents.

The candidate formulae obtained from MassWorks, and the odd molecular ion, both indicated the presence of an odd number of nitrogen atoms [6]. An IR spectrum was provided for the unknown [25] to assist in identifying the nitrogen functional group as an amide. Having IR spectral information available was considered reasonable because this would be part of the routine analysis of an unknown in a forensic laboratory. Formulae of high spectral match quality were evaluated for their ability to generate the formulae obtained for fragment ions

Table 7b. Unknown A42 (-H mixture search).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C8H19N2O3	191.1396	5.8675	30.6983	99.2152	7	0.5
C7H17N3O3	191.1270	−6.7086	−35.0988	99.1928	7	1.0
C10H15N4	191.1297	−4.0285	−21.0769	99.1790	7	5.5
C9H19O4	191.1283	−5.3659	−28.0741	99.1591	7	0.5
C11H17N3	191.1422	8.5476	44.7203	99.0932	8	5.0
C12H17NO	*191.1310*	*−2.6858*	*−14.0521*	*98.9401*	*9*	*5.0*
C13H19O	191.1436	9.8902	51.7451	98.8126	11	4.5
C6H19N4OSi	191.1328	−0.8872	−4.6419	96.3572	31	0.5
C7H19N4S	191.1330	−0.6577	−3.4409	95.9274	33	0.5
C7H21N3OSi	191.1454	11.6888	61.1553	95.8778	34	0.0

Note: Italics show the correct formula for the target molecule.

Table 7c. Unknown A42 (-H mixture search, odd ions only).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C7H17N3O3	191.1270	−6.7086	−35.0988	99.1928	7	1.0
C11H17N3	191.1422	8.5476	44.7203	99.0932	8	5.0
C12H17NO	*191.1310*	*−2.6858*	*−14.0521*	*98.9401*	*9*	*5.0*
C7H21N3OSi	191.1454	11.6888	61.1553	95.8778	34	0.0
C8H21N3S	191.1456	11.9184	62.3562	95.6591	35	0.0
C8H21NO2Si	191.1342	0.4555	2.3829	95.4893	36	0.0
C9H21NOS	191.1344	0.6850	3.5838	95.3084	37	0.0
C6H21N3Si2	191.1274	−6.2992	−32.9573	91.5766	66	0.0
C8H18N3Cl	191.1189	−14.7747	−77.3003	67.3500	253	1.0
C10H22NCl	191.1441	10.3774	54.2940	66.8927	257	0.0

Note: Italics show the correct formula for the target molecule.

using an iterative process. Only one formula in the top three formula candidates (Table 7c) could produce the proposed formulae for all four of the fragment ions in Table 7d.

It was clear from the fragment formulae that a toluene moiety was present and was generated by a loss of carbon monoxide. Furthermore, the fragment that produced the toluene fragment resulted from the loss of the lone nitrogen present (indicated by a shift from even mass fragments to odd ones [6]) and four sp^3 carbons. Since a forensic scientist may encounter any substance in existence, each of the nine prospective structures shown in the schematic in Figure 6 could represent the correct compound. However, forensic assessments must also account for the probability of encountering any given substance, where the most likely compound is the one that is most readily available to the public. A search was performed using NIST Chemistry WebBook against the molecular formula capable of generating all of the fragment ions, comparing the unknown mass spectrum to the reference mass spectrum for each structural isomer. In this case, n,n-diethyl metatoluamide, the insect repellent DEET, is the most readily available and is supported by the assignment of the fragmentation mass spectrum provided in Figure 7. Had no suitable match been present within the NIST Chemistry WebBook for this or any of the other unknowns, PubChem and ChemSpider would have been consulted, with a higher probability given to database matches with a higher number of citations, in accordance with the rationale presented by Little, et al. [22]. This unknown compound was correctly identified as DEET.

Table 7d. Unknown A42 (most reasonable fragment formulae).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C8H7O	119.0497	−1.8102	−15.2048	98.0954	40	5.5
C7H5NO	119.0371	−14.3862	−120.8402	98.0577	42	6.0
C7H7	91.0548	6.5752	72.2170	95.0432	38	4.5
C6H5N	91.0422	−6.0008	−65.9083	94.7788	41	5.0

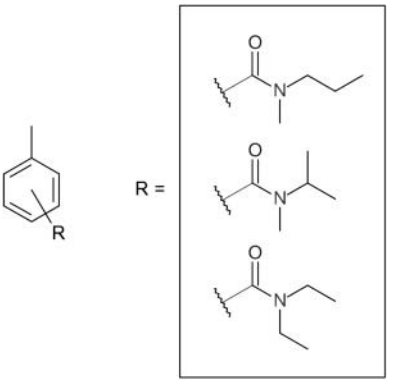

Figure 6. Schematic for the nine most likely structural isomers for Unknown A42, determined from CLIPS searches of fragment ions and evaluations of the neutral losses giving rise to those fragments.

Blind test 2: mixture of three unknown compounds

Given that unknown compound A42 (DEET) provided a well-defined mass spectrum with little to no ion interferences, it was considered to be a simple case for evaluating MassWorks. Another single-blind experiment, a mixture of three unknowns, was used to further test MassWorks. One unknown had an easily identified molecular ion with strong, well separated fragments (Peak 1); the second

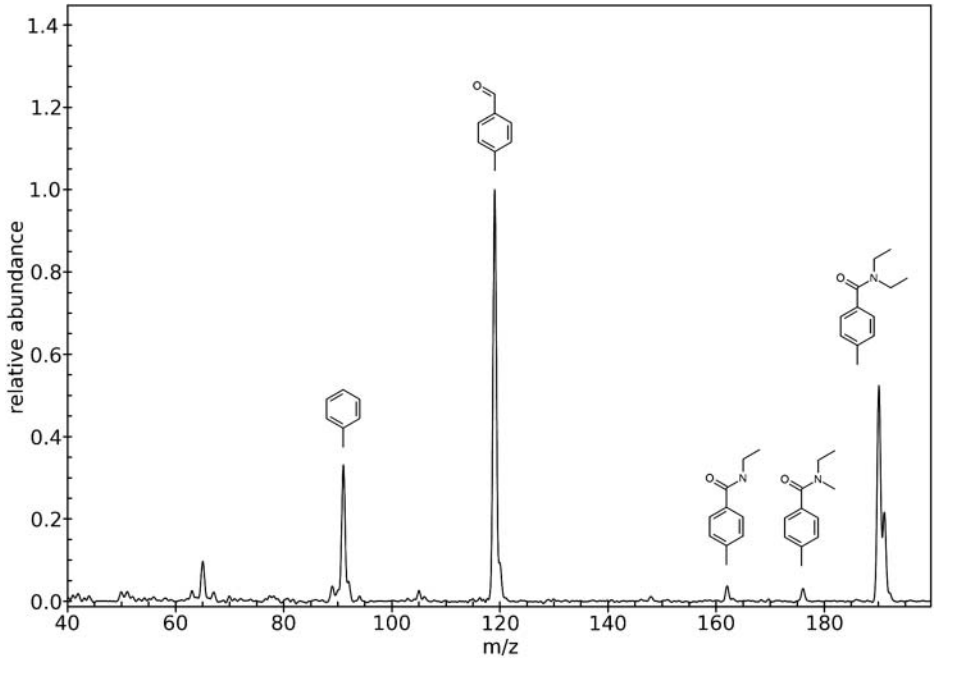

Figure 7. Mass spectrum of Unknown A42, with structural fragments illustrated for DEET (other possible structures for each fragment were also considered in identifying the unknown).

had a strong molecular ion with significantly overlapped, low abundance fragment ions (Peak 2); and the third had strong, but overlapped fragment ions with no molecular ion (Peak 3). These unknowns were provided as a single clear, colourless methanol solution. The three unknowns eluted with retention times of 7.4, 8.9, and 18.2 min respectively.

Peak 1 chromatographed poorly, and by stepping through the peak one scan at a time, it was apparent that there was some degradation of the compound apart from the ionization process because the ion abundances, and even the presence or absence of ions, was not consistent throughout the peak. The heaviest identifiable fragment ion for Peak 1 was 118.0740 m/z with a pattern implying the loss of one hydrogen, and a subsequent loss of another two. Higher ion masses were present, but of inadequate intensity to be attributed to the unknown instead of column bleed or the coelution of an impurity. Reducing the ionization energy to 11 eV was attempted, to see if the molecular ion would become visible, to no avail. A CLIPS search was performed on the 115, 117 and 118 m/z ions using the mixture search fields to account for the appropriate losses or additions of hydrogens in each case, with the best result achieved for the fragment at 115 m/z. A CLIPS search was also run on fragments 105, 91, 77 and 58, suggesting the presence of either a butane or aminopropane structural component and the presence of an aromatic ring. In the absence of better information, it was not possible to assign a molecular formula to this compound with the aid of MassWorks alone. This illustrated how, in a real-world situation, searching the mass spectrum against a database (e.g. NIST 02) may have provided the information necessary to identify the compound using the MassWorks software. Another approach may have been to chemically modify the analyte in an attempt to impart higher thermal stability to the compound or lower the instability of the molecular ion. On concluding that an identification could not be reached, the identity of the unknown, pseudoephedrine, was disclosed: a classic example of a compound with atypical fragmentation, and an unobserved molecular ion under standard EI conditions [26].

The heaviest ion for Peak 2 was a well-defined peak at 165.0860 m/z. This peak was accompanied by a small hump one m/z unit below the monoisotopic peak, implying the presence of a low abundance single hydrogen loss fragment ion. Including the relative formula '-H' in the mixture field and specifying ODD ions only for the CLIPS search returned a list of candidates (Table 8a) with the first candidate having an excellent spectral fit based on both the spectral accuracy value provided by MassWorks, and the visual similarity between the observed and theoretical isotopic shapes (Figure 8). Following the same iterative process used for unknown compound A42, the top two candidate formulae from the list in Table 8b were evaluated for their ability to generate the observed fragment ion formulae for ions 137, 120, 92/91 and 65 m/z. The presence of the structural components ethoxide, a loss of carbon monoxide, and either an aniline or methylpyridine ring were identified, which are most logically derived from the highest ranked molecular formula, $C_9H_{11}NO_2$. The structural isomers that can be generated from the fragment formulae are presented by the schematic in Figure 9. The proposed molecular formula was searched in NIST Chemistry WebBook, and the search results manually filtered to exclude any candidates lacking the structural features determined from the fragment ions. Of the remaining compounds, ethyl para-aminobenzoate, the topical anaesthetic benzocaine, is the most readily available to the public, and was supported by the assignment of the fragmentation

Table 8a. Peak 2 (-H mixture search, odd ions only).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C9H11NO2	*165.0790*	*−7.0214*	*−42.5318*	*99.4659*	*823*	*5.0*
C8H11N3O	165.0902	4.2120	25.5139	99.3687	988	5.0
C4H15N3O2Si	165.0934	7.3533	44.5421	98.0999	2820	0.0
C5H15NO3Si	165.0821	−3.8801	−23.5036	98.0456	2880	0.0
C7H11N3Si	165.0722	−13.7761	−83.4480	97.6971	3370	5.0
C6H15NO4	165.1001	14.1080	85.4583	97.5524	3769	0.0
C4H11N3O4	165.0750	−11.0441	−66.8993	96.5233	5354	1.0
C9H15NSi	165.0974	11.3760	68.9097	96.2844	5224	4.0
C6H15NO2S	165.0823	−3.6506	−22.1132	95.5943	6234	0.0
C5H15N3OS	165.0936	7.5828	45.9325	95.4942	6548	0.0

Note: Italics show the correct formula for the target molecule.

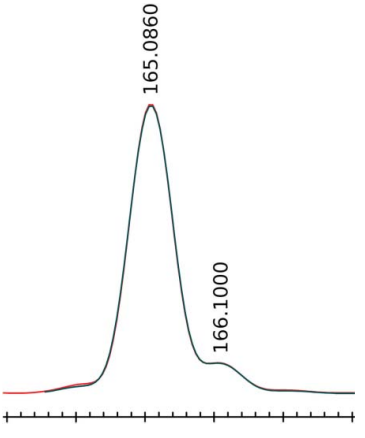

Figure 8. CLIPS spectral fit for Peak 2 of the Unknown mixture for $C_9H_{11}NO_2$ with a mixture formula of –H.

Table 8b. Peak 2 (most reasonable fragment formulae).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C7H7NO2	137.0477	1.5785	11.5178	99.1027	721	5.0
C8H9O2	137.0603	14.1545	103.2830	98.8501	917	4.5
C8H8O	120.0575	10.3149	85.9235	97.9465	8890	5.0
C7H6NO	120.0449	−2.2612	−18.8358	97.5390	10741	5.5
C6H6N	92.0500	7.5242	81.7470	99.2282	990	4.5
C4H3N	65.0265	−6.7509	−103.8068	99.6787	388	4.0
C5H5	65.0391	5.8252	89.5720	99.6563	411	3.5

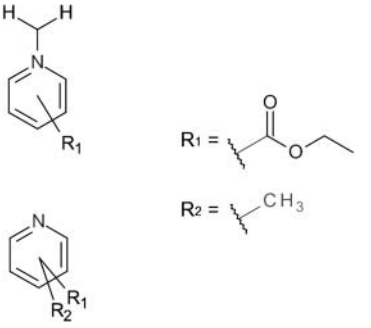

Figure 9. Schematic for the most likely structural isomers for Peak 2 of the Unknown mixture, determined from CLIPS searches of fragment ions and evaluations of the neutral losses giving rise to those fragments.

mass spectrum (Figure 10). This unknown compound was correctly identified as benzocaine.

Peak 3 presented a well-defined molecular ion at 334.1670 m/z, its shape implying ion overlap from the loss of a single hydrogen. Unlike the two previous unknowns, only one clearly defined fragment ion free of spectral interference was present (319 m/z). A CLIPS search was performed on the molecular ion, including the relative formula '-H' in the mixture field and specifying ODD ions only,

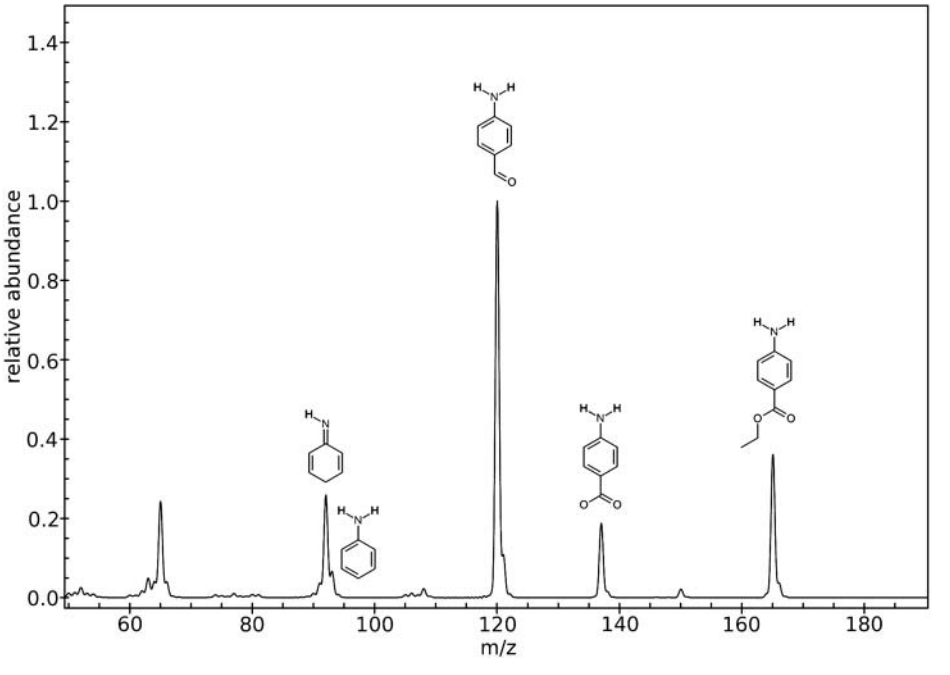

Figure 10. Mass spectrum of Peak 2 of the Unknown mixture, with structural fragments illustrated for Benzocaine (as with DEET, other possible structures for each fragment were also considered in identifying the unknown).

Table 9a. Peak 3 (-H mixture search, odd ions only).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C20H22N4O	334.1794	11.9613	35.7945	98.8144	266	12.0
C21H22N2O2	*334.1681*	*0.7280*	*2.1784*	*98.6839*	*293*	*12.0*
C22H22O3	334.1569	−10.5054	−31.4376	98.5314	312	12.0
C19H26O5	334.1780	10.6239	31.7923	97.9998	434	7.0
C17H22N2O5	334.1529	−14.5282	−43.4757	97.2459	614	8.0
C16H22N4O4	334.1641	−3.2948	−9.8597	97.1360	643	8.0
C17H26N2O3Si	334.1713	3.8692	11.5788	97.1308	585	7.0
C18H26O4Si	334.1600	−7.3641	−22.0373	96.9260	626	7.0
C15H30O6Si	334.1812	13.7652	41.1926	96.5986	693	2.0
C23H26S	334.1755	8.1215	24.3038	96.3707	735	11.0

Note: Italics show the correct formula for the target molecule.

to generate a list of candidates (Table 9a). In an attempt to assess the highly overlapped fragment ions, mixture formulae were defined with losses or additions of hydrogen to bracket the peak of interest in a qualitative fashion for fragments 319, 306, 277, 263, 168, 163, 162, 161, 144, 143, 130, 115, 108, 107, 91 and 77; however, this approach yielded little more than the suggestion of the presence of

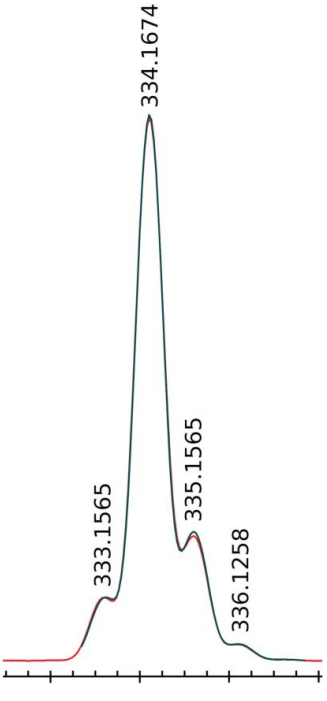

Figure 11. CLIPS spectral fit for Peak 3 of the Unknown mixture for $C_{21}H_{22}N_2O_2$ with a mixture formula of –H.

an aromatic ring and a large series of fragment formulae (Table 9b) that could be compared with the candidate molecular ion formulae. A clear pattern of repeating -CH_2 losses was evident immediately below the molecular ion and, in conjunction with the calculated value of 12 double bond equivalents for the best three matches, implied a highly polycyclic structure. Evaluation of the fragment list for formulae that would support any of the top three proposed molecular ion formulae

Table 9b. Peak 3 (most reasonable fragment formulae).

Formula	Mono isotope (m/z)	Mass error (mDa)	Mass error (PPM)	Spectral accuracy (%)	RMSE	DBE
C20H19N2O2	319.1447	0.7529	2.3590	98.9824	14	12.5
C21H21NO2	319.1572	13.3289	41.7646	98.6706	18	12.0
C19H18N2O2	306.1368	7.3278	23.9370	97.3063	28	12.0
C17H13N2O2	277.0977	−3.8973	−14.0646	86.5187	128	12.5
C18H15NO2	277.1103	8.6787	31.3197	86.0110	132	12.0
C18H17NO	263.1310	10.2142	38.8193	86.3946	136	11.0
C17H15N2O	263.1184	−2.3619	−8.9764	86.3837	136	11.5
C11H8N2	168.0687	−6.9517	−41.3607	92.3484	192	9.0
C12H10N	168.0813	5.6243	33.4631	92.3481	191	8.5
C8H12N2O2	168.0899	14.1776	84.3527	91.9908	200	4.0
C9H11N2O	163.0871	−12.8620	−78.8597	93.7294	145	5.5
C10H13NO	163.0997	−0.2860	−1.7533	93.6728	145	5.0
C11H15O	163.1123	12.2901	75.3532	93.6099	146	4.5
C11H14O	162.1045	12.9651	79.9861	97.5873	83	5.0
C10H12NO	162.0919	0.3890	2.4000	97.5115	85	5.5
C9H10N2O	162.0793	−12.1870	−75.1862	97.4329	89	6.0
C10H11NO	161.0841	−3.1360	−19.4678	98.0538	74	6.0
C11H13O	161.0966	9.4400	58.6020	98.0428	74	5.5
C9H8N2	144.0687	−5.0517	−35.0635	89.3738	376	7.0
C10H10N	144.0813	7.5243	52.2255	89.1920	383	6.5
C6H11N2O2	143.0821	8.3526	58.3798	94.1963	239	2.5
C9H7N2	143.0609	−12.7768	−89.3020	93.8945	255	7.5
C10H9N	143.0735	−0.2007	−1.4028	93.7401	260	7.0
C11H11	143.0861	12.3754	86.4964	93.5723	265	6.5
C8H6N2	130.0531	−4.4018	−33.8450	88.2084	521	7.0
C9H8N	130.0657	8.1743	62.8511	87.7520	538	6.5
C9H7	115.0548	8.1752	71.0601	90.4021	226	6.5
C8H5N	115.0422	−4.4008	−38.2526	89.9224	237	7.0
C4H7N2O2	115.0508	4.1525	36.0939	88.4775	271	2.5
C7H10N	108.0813	0.6243	5.7765	90.9527	217	3.5
C6H8N2	108.0687	−11.9517	−110.5816	90.8939	221	4.0
C8H12	108.0939	13.2004	122.1345	90.8789	212	3.0
C7H7O	107.0497	−13.2102	−123.3868	88.8518	331	4.5
C7H9N	107.0735	10.5993	99.0006	88.7271	353	4.0
C6H7N2	107.0609	−1.9768	−18.4636	88.7220	356	4.5
C7H7	91.0548	9.8752	108.4654	92.7159	138	4.5
C6H5N	91.0422	−2.7008	−29.6649	92.6194	140	5.0
C2H7N2O2	91.0508	5.8525	64.2812	91.5488	162	0.5
CH5N2O2	77.0351	10.0024	129.8592	88.0445	308	0.5
C6H5	77.0391	14.0252	182.0856	87.4970	287	4.5
C5H3N	77.0265	1.4491	18.8134	87.1773	294	5.0
C4HN2	77.0140	−11.1270	−144.4589	86.7032	304	5.5

resulted in the third formula ($C_{22}H_{22}O_3$) being excluded due to lack of supporting fragments, and the first candidate ($C_{20}H_{22}N_4O$) deemed less likely than the second due to a lower number of supporting fragment formulae than for $C_{21}H_{22}N_2O_2$. When taking the mass error into consideration, the formula $C_{21}H_{22}N_2O_2$ exhibited the lowest mass error of the three candidates with a good spectral fit (Table 9a and Figure 11). A search of the NIST Chemistry WebBook provided 'strychnine' as the only recorded molecule with this molecular formula, the identification of which was supported by a comparison with a reference mass spectrum (Figure 12). This unknown compound was correctly identified as strychnine.

Conclusion

The MassWorks software uses a novel approach for achieving significantly increased mass accuracy from the common, single quadrupole mass spectrometer. This software has been shown to significantly increase the speed and simplicity of determining the molecular formula, structural components and ultimately the identity of an unknown compound. In addition to improving the calibration of quadrupole data, the spectral fitting of isotopic peaks allows the software to evaluate the quality of proposed elemental formulae without the need for complicated filtering rules. This, in turn, allows shorter candidate lists to be submitted to online databases in identifying an unknown. Finally, with the ability to determine

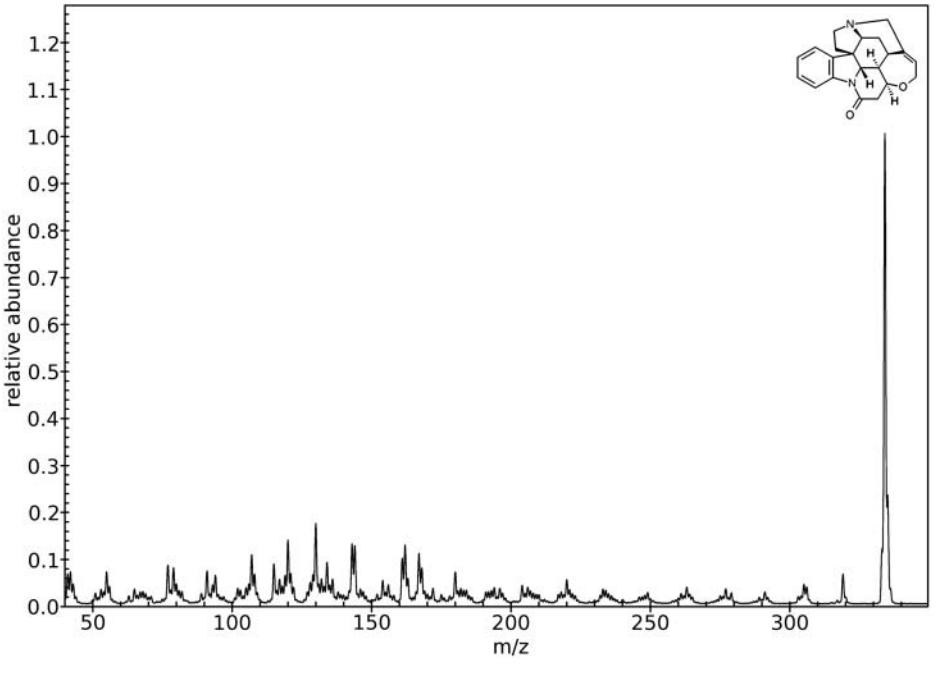

Figure 12. Mass spectrum of Peak 3 of the Unknown mixture, with the only structural match to $C_{21}H_{22}N_2O_2$ within NIST Chemistry WebBook marking the molecular ion. Fragment formulae (see Table 9b) were considered in arriving at the correct molecular formula, but not used for assessing structure.

molecular formulae for fragment ions using a standard GC-EI-MS instrument, MassWorks permits comprehensive unknown identification without the need for complicated or expensive instrumentation.

Acknowledgement

The author would like to acknowledge the generous technical and editorial support offered throughout this study by Gene Basara and Mark Sandercock.

References

1. Capurro PU. Effects of chronic exposure to solvents caused by air pollution. Clin Toxicol. 1970;3(2):233–248.
2. Mrochek JE, Rainey WT, Jr. Identification and biochemical significance of substituted furans in human urine. Clin Chem. 1972;18(8):821–828.
3. Castorena JL, Garriott JC, Barnhardt FE, Shaw RF. A fatal poisoning from Nicotiana glauca. J Toxicol-Clin Toxicol. 1987;25(5):429–435.
4. Skoog DA, Holler FJ, Nieman TA. Principles of instrumental analysis. 5th. Orlando (FL): Saunders College Pub.; Harcourt Brace College Publishers; 1998.
5. Liao W, Draper WM, Perera SK. Identification of unknowns in atmospheric pressure ionization mass spectrometry using a mass to structure search engine. Anal Chem. 2008;80(20):7765–7777.
6. McLafferty F, Tureček F. Interpretation of mass spectra. 4th. Mill Valley (CA): University Science Books; 1993.
7. Finkle BS, Taylor DM, Bonelli EJ. A GC-MS reference data system for the identification of drugs of abuse. J Chromatogr Sci. 1972;10(5):312–333.
8. Wang Y, Gu M. The concept of spectral accuracy for MS. Anal Chem. 2010;82 (17):7055–7062.
9. 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.
10. Erve JC, Gu M, Wang Y, DeMaio W, Talaat RE. Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination. J Am Soc Mass Spectr. 2009;20(11):2058–2069.
11. Eckenrode BA. Environmental and forensic applications of field-portable GC-MS: an overview. J Am Soc Mass Spectr. 2001;12(6):683–693.
12. Wang Y, Prest H. MassWorks User's Guide and Quick Start Tutorial. Danbury, CT: Cerno Bioscience LLC; 2006. User Manual.
13. Linstrom PJ, Mallard WG. NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Gaithersburg (MD): National Institute of Standards and Technology [Internet]. Available from: http://webbook.nist.gov.
14. Kovats E. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. Helv Chim Acta. 1958;41:1915–1932.
15. van Den Dool H, Dec. Kratz P. A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography. J Chromatogr A. 1963;11(0):463–471.
16. de Zeeuw RA. Substance identification: the weak link in analytical toxicology. J Chromatogr B. 2004;811(1):3–12.
17. Martz RM, Reutter DJ, Lasswell LD, III. A Comparison of Ionization Techniques for Gas Chromatography/Mass Spectroscopy Analysis of Dye and Lachrymator Residues from Exploding Bank Security Devices. J Forensic Sci. 1983;28(1):200–207.
18. McNair HM, Miller JM. Basic gas chromatography. New York (NY): Wiley; 1998.
19. Bolton EE, Wang Y, Thiessen PA, Bryant SH, Ralph AW, David CS. Annual Reports in Computational Chemistry. Elsevier; 2008. Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities; Computational Chemistry.

20. Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. J Chem Educ. 2010;87(11):1123–1124.
21. Kind T, Fiehn O. Advances in structure elucidation of small molecules using mass spectrometry. Bioanal Rev. 2010;2(1–4):23–60.
22. Little JL, Cleven CD, Brown SD. Identification of "known unknowns" utilizing accurate mass data and chemical abstracts service databases. J Am Soc Mass Spectr. 2011;22(2):348–359.
23. Stein SE. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J Am Soc Mass Spectr. 1999;10 (8):770–781.
24. Kind T, Fiehn O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 2007;8:105.
25. Kinugasa S, Tanabe K, Tamura T, Wasada N, Yamaji T, Saito T, Hayamizu K, Yanagisawa M, Yamamoto O, Someno K, et al. Spectral Database for Organic Compounds, SDBS [Internet]. Available from: http://sdbs.db.aist.go.jp
26. Smith RM, Busch KL. Understanding mass spectra: a basic approach. New York (NY): Wiley; 1999.