

UC Berkeley

UC Berkeley Previously Published Works

Title

Improved molecular level identification of organic compounds using comprehensive two-dimensional chromatography, dual ionization energies and high resolution mass spectrometry

Permalink

<https://escholarship.org/uc/item/4n6156zd>

Journal

Analyst, 142(13)

ISSN

0003-2654

Authors

Worton, David R
Decker, Monika
Isaacman-VanWertz, Gabriel
et al.

Publication Date

2017-07-07

DOI

10.1039/c7an00625j

Peer reviewed

Improved molecular level identification of organic compounds using comprehensive two-dimensional chromatography, dual ionization energies and high resolution mass spectrometry

David R. Worton^{1,&*}, Monika Decker¹, Gabriel Isaacman-VanWertz^{1,§}, Arthur W. H. Chan^{1,#}, Kevin R. Wilson² and Allen H. Goldstein^{1,3}

¹ Department of Environmental Sciences Policy and Management, University of California, Berkeley, CA, 94720, USA.

² Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

³ Department of Civil and Environmental Engineering, University of California, Berkeley, CA, 94720, USA.

[&] Now at National Physical Laboratory, Hampton Road, Teddington, Middlesex TW11 0LW, UK.

[§] now at Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, 24061.

[#] now at Department of Chemical Engineering and Applied Chemistry, University of Toronto, Ontario, M5S 3E5, Canada.

*Corresponding author: dave.worton@npl.co.uk, Tel. +44 208943 6591.

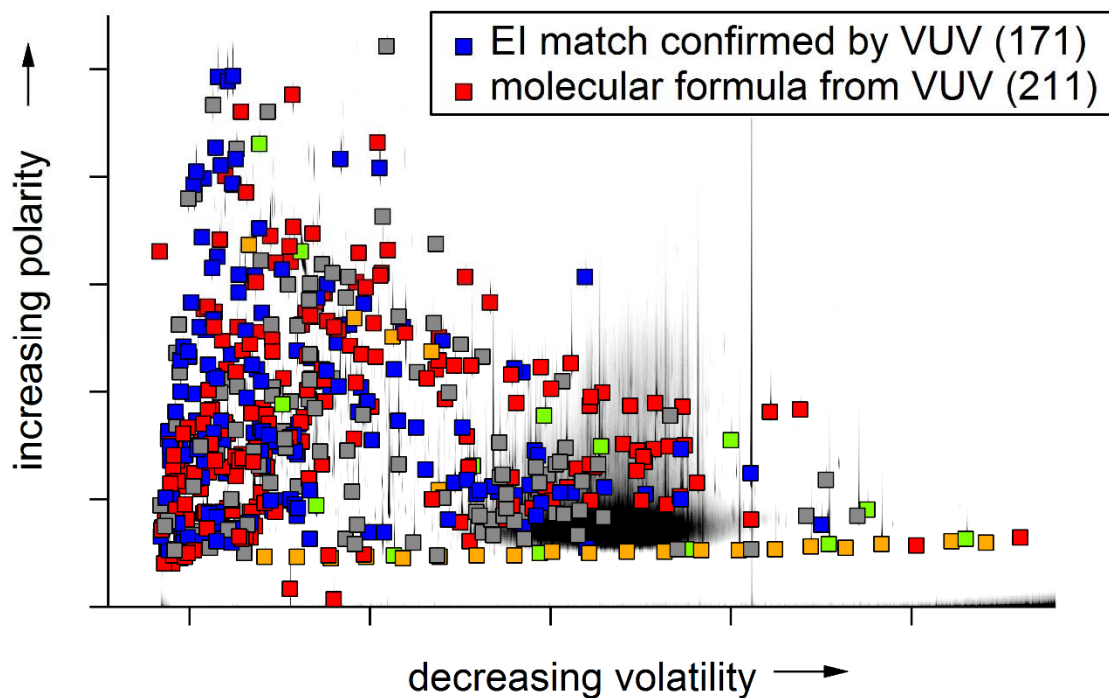
Abstract

A new analytical methodology combining comprehensive two-dimensional gas chromatography (GC×GC), dual ionization energies and high resolution time of flight mass spectrometry has been developed that improves molecular level identification of organic compounds in complex environmental samples. GC×GC maximizes compound separation providing cleaner mass spectra by minimizing erroneous fragments from interferences and co-eluting peaks. Traditional electron ionization (EI, 70 eV) provides MS fragmentation patterns that can be matched to published EI MS libraries while vacuum ultraviolet photoionization (VUV, 10.5 eV) yields MS with reduced fragmentation enhancing the abundance of the molecular ion providing molecular formulas when combined with high resolution mass spectrometry. We demonstrate this new approach by applying it to a sample of organic aerosol. In this sample, 238 peaks were matched to EI MS library data with FM \geq 800 but a fifth (42 compounds) were determined to be incorrectly identified because the molecular formula was not confirmed by the VUV MS data. This highlights the importance of using a complementary technique to confirm compound identifications even for peaks with very good matching statistics. In total, 171 compounds were identified by EI MS matching to library spectra with confirmation of the molecular formula from the high resolution VUV MS data and were not dependent on the matching statistics being above a threshold value. A large number of unidentified peaks were still observed with FM < 800, which in routine analysis would typically be neglected. Where possible, these peaks were assigned molecular formulas from the VUV MS data (211 in total). In total, the combination of EI and VUV MS data provides more than twice as much molecular level peak information

than traditional approaches and improves confidence in the identification of individual organic compounds. The molecular formula data from the VUV MS data was used, in conjunction with GC×GC retention times and the observed EI MS, to generate a new, searchable EI MS library compatible with the standard NIST MS search program. This library is deliberately dynamic and editable so that other end users can add new entries and update existing entries as new information becomes available.

TOC abstract and statement

A new analytical methodology has been developed to improve molecular level identification of organic compounds in complex samples.



INTRODUCTION

Organic compounds are ubiquitous in the Earth's biogeochemical reservoirs and span a wide range of carbon chain lengths and functional group classes.¹ Substantial quantities of organic compounds are emitted into the environment by a variety of natural and man-made sources, e.g., industrial emissions, combustion sources and the terrestrial biosphere.² Once in the environment these compounds can be transformed through chemical and biological oxidation processes leading to the formation of oxygenated species that dramatically increases the complexity of the organic mixture.^{1,3,4} Chemical speciation of individual organic 'tracer' compounds in complex environmental mixtures provides critical information for apportioning organic carbon from an often convoluted mixture of sources and for constraining transformational processes critical for environmental fate modeling.⁵⁻⁸ However, the vast majority of organic compounds present in the environment remain unidentified and missing from published electron ionization mass spectral (EI MS) libraries due to difficulties in separating and identifying them. As a result, important and potentially unique 'tracer' compounds remain uncharacterized, and much of the organic mass remains part of an unresolved complex mixture, using classically available analytical techniques.

Methods for identification of individual compounds have generally relied on matching gas chromatography (GC) retention times and MS fragmentation patterns to authentic standards and published mass spectral libraries of known compounds (e.g., NIST).⁹ However, even with high chromatographic resolution, peak co-elution is common due to sample complexity. This can lead to ambiguous mass spectra, which complicates interpretation and reduces confidence in compound identification. This issue is largely mitigated by using two dimensional GC (GC×GC), which is now a well-established technique that is becoming more widely used.^{10,11} First introduced by Liu and Phillips¹², GC×GC couples two chromatographic columns with different stationary phases in series by way of a modulator that traps effluent from the first column and rapidly injects it onto the second.¹³ This provides higher peak resolution, higher peak capacity and greater selectivity than traditional one dimension GC at the expense of more complex data requiring more sophisticated data analysis tools and more skilled operators. Minimal peak co-elution facilitates the acquisition of 'cleaner' mass spectra that are free of significant interferences.^{12,14} Nevertheless, the range of observed organic compounds far exceeds the number of compounds available as authentic standards or published in EI MS libraries. This is especially true for oxygenated products, many of which may be uniquely created by chemical and/or biological processes and may not be sufficiently stable to synthesize and store. Identification of a broader range of organic compounds is critical to further improve our understanding of the fate and potential impacts of organic compounds in the environment so improved analytical methodologies for chemical identification are needed.

In this work, we combine GC×GC, traditional 70 eV EI and vacuum ultraviolet (VUV) photoionization with high resolution time of flight mass spectrometry providing a new approach to better characterize unidentified compounds in complex environmental mixtures. EI provides mass spectral fragmentation patterns that can be matched to published EI MS libraries as well as used to elucidate structures for unknown

compounds. VUV photoionization is a ‘soft’ technique that reduces fragmentation and enhances the abundance of the molecular ion. When combined with high resolution mass spectrometry, this provides molecular formulas for chromatographic peaks, which can be used for identification purposes. VUV single photon ionization mass spectrometry (SPI-MS) has proven to be a valuable tool for the analysis of complex mixtures because the molecular ions can be used for identification purposes.^{15, 16} The energy of the VUV photons can be tuned to be just above the ionization energy of the target analyte minimizing fragmentation of the molecular ion allowing the parent mass peak to dominate the mass spectrum. For most purposes 10.5 eV has been determined to be a useful energy level.¹⁷⁻¹⁹ High resolution mass spectrometry also facilitates grouping of the observed chromatographic peaks by their Kendrick mass defect (CH₂) providing a mechanism to identify homologous series.²⁰ The EI and VUV ionization schemes are complementary and together represent a unique approach for the identification of unknown organic compounds. In this work, we demonstrate this approach by applying it to identify individual organic compounds in atmospheric particulate matter and to develop a new EI MS library composed of novel, atmospherically relevant compounds mass spectra that can be subsequently utilized, similar to other published EI MS libraries for sample analysis, by anyone using traditional EI. Our methodology could be applied to a wide variety of different samples across a broad spectrum of research fields generating additional entries to existing EI MS libraries or for the generation of new EI MS libraries that could be made publically available to the research community.

EXPERIMENTAL

Field Site and Sample Collection. High volume filter samples (26 in total) of atmospheric particulate matter with aerodynamic diameters of less than 2.5 microns (PM_{2.5}) were collected at the University of California’s Blodgett Forest Research Station (UC-BFRS, 38.90 °N, 120.63 °W, 1315 m elevation above sea level) in July 2009 as part of the Biosphere Effects on AeRosols and Photochemistry Experiment (BEARPEX).²¹ UC-BFRS is located between Sacramento and Lake Tahoe in the Sierra Nevada Mountains and was established as an atmospheric measurement site in 1997.²² Air sampled at this site is influenced by local biogenic emissions from mixed conifer forests dominated by ponderosa pine trees, aged biogenic emissions from oak trees located several hours upwind and urban outflow from the Sacramento region.^{22, 23} Samples were collected at 1130 lpm for 6 – 12 hours (sample volumes of 30 – 80 m³ of air) on quartz filters (20 × 25 cm, Gellman QM-A) that had been baked at 600 °C for 12 hours and stored in pre-baked aluminum foil prior to use. After sampling the quartz filters were immediately placed in pre-baked aluminum foil and stores at -20 C until analysis. Typical aerosol loadings during sample collection were ~ 5 µg m⁻³ with ~75 % organics by mass.²¹

Instrumentation and data analysis. Portions of the filters (punches of 0.8 cm² were made using an Arch Punch; McMaster Carr) were thermally desorbed at 320°C under a helium flow using a thermal desorption system and autosampler (TDS3 and TDSA2, Gerstel). Desorbed samples were focused at 10°C on a quartz wool liner in a cooled injection system (CIS4, Gerstel) prior to introduction into the gas chromatograph (GC, Agilent 7890) by rapidly heating the inlet to 320°C. Comprehensive GC×GC was performed

using a $60\text{ m} \times 0.25\text{ mm} \times 0.25\mu\text{m}$ nonpolar capillary column (Rxi-5Sil MS, Restek) for the first dimension separation and a medium-polarity second dimension column ($1\text{ m} \times 0.25\text{ mm} \times 0.25\mu\text{m}$, Rtx-200MS, Restek). A dual-stage thermal modulator (Zoex), consisting of a guard column ($1\text{ m} \times 0.25\text{ mm}$, Rxi, Restek), was used as the interface between the two columns, where the effluent from the first column was cryogenically focused and periodically injected into the second column every 2.3 seconds. The second dimension column was housed in a secondary oven, which was maintained at 15°C above the main oven temperature. The higher temperature of the second oven relative to the first reduced retention of analytes in the second column ensuring they all eluted within 2.3 seconds of injection preventing undesirable peak ‘wrap around’ that can complicate data analysis. The carrier gas flow rate was 2 ml/min of helium and the GC temperature program was 40°C hold for 2 min, $3.5^\circ\text{C}/\text{min}$ until 320°C and hold for 10 min.

Analytes eluting from the second column were ionized using 70 eV EI or 10.5 eV VUV always using a high resolution ($m/\Delta m \sim 4000$) time of flight mass spectrometer (HTOF, Tofwerk). All filter samples were analyzed twice, once using EI and once with VUV. To minimize fragmentation in VUV the ion source was operated at a lower temperature of 170°C relative to EI (270°C). Photons of 10.5 eV with an energy distribution width of 0.2 eV and an intensity of $\sim 10^{15}$ photons s^{-1} were generated by the Chemical Dynamics Beamline (9.0.2) of the Advanced Light Source at the Lawrence Berkeley National Laboratory.^{18, 24} The photon beam was introduced to the ionization chamber perpendicular to the column outflow and at a 45° angle to the EI filaments facilitating photoionization without removal of the EI filaments. Since the ionization energies of most organic compounds lie between 8 and 11 eV, the minimal excess energy limits fragmentation of ionized molecules allowing for substantial higher abundances of the molecular ions relative to traditional EI. The frequency of data collection was 200 Hz to provide sufficient datapoints needed to adequately reconstruct the narrow chromatographic peaks ($\sim 100\text{ ms}$) eluting from the secondary column.

The molecular ions were determined to be the largest observed m/Q above a specified threshold, defined as 1 % of the total ion signal in this work. The unit resolution molecular weight had to be even unless an odd number of nitrogen atoms were present. To ensure comparability between different research groups, retention times in this work are reported as Kovats retention indices (I_1 ; first dimension retention times)²⁵ and Pankow retention time ratio (R_2 ; second dimension retention times).²⁶ In both cases, deuterated n-alkanes were used as the reference compounds with the I_1 of the deuterated n-alkanes having been determined during this work relative to authentic standards of n-alkanes through co-injections of known standards. The Kendrick mass defect (KMD) was calculated as the difference between the nominal mass and the observed high resolution after converting the Kendrick mass using a mass of 14.0000 for CH_2 .²⁰

All data processing was performed using custom data analysis code written in Igor 6.37 (Wavemetrics)^{17, 18} based on techniques developed for the analysis of high resolution mass spectral data collected by an Aerodyne Aerosol Mass Spectrometer (AMS) equipped with a comparable mass spectrometer.²⁷ The spectrometric signal at a given nominal mass was deconvoluted into component peaks by minimizing the residual of a fit

based on exact masses of candidate peaks likely to be present. For the EI high resolution data the candidate peaks used were all the carbonaceous fragments reported in the AMS peak integration by key analysis (PIKA) fragmentation tables.²⁷ For VUV high resolution data, a custom fragmentation table was determined.¹⁷ To account for deviations from Gaussian peak shapes and drifts in mass calibration, a peak shape was selected and mass calibration was performed using frequently and regularly spaced per-deuterated normal alkanes, spiked onto the filter before desorption, that covered the volatility range of the observed analytes (C₁₂ – C₂₈; C/D/N Isotopes). The custom IGOR data analysis code links directly with the widely used NIST mass spectrometer database search routine developed by Stein.^{9, 28} In this work we use version 2.0g of the NIST search program and the 2011 version of the NIST mass spectral library. Detection of chromatographic peak locations was carried out using GC Image software (LLC).

RESULTS AND DISCUSSION

Figure 1a shows a representative GC×GC chromatogram from the analysis of a filter sample collected during the BEARPEX campaign with more than 500 separated peaks illustrating the inherent organic chemical complexity in atmospheric aerosol samples. Figure 1b shows the same GC×GC chromatogram but replotted against I₁ and R₂ along with the classification of observed peaks that will be discussed in this paper indicated by the colored symbols. As samples were analyzed with both EI and VUV ionization, every peak has two associated mass spectra (MS), one EI and one VUV. The advantage of GC×GC over conventional one-dimensional GC is that it substantially reduces co-elutions leading to ‘cleaner’ spectra with minimal interferences, thus providing better matches to library spectra (EI)¹⁴ and more confidence in the identification of the molecular ion (VUV).

Retention time matching, with confirmation by mass spectral matching, to authentic standards is the most accurate way to identify chromatographic peaks in unknown samples. In this work, an authentic standard mixture containing 96 individual compounds spanning a range of carbon numbers (C₇ – C₃₆) and functional groups (alkanes, alkenes, PAHs, aldehydes, ketones, monocarboxylic acids) was used but only 25 of these compounds were positively identified. Only using these severely limits the obtainable information such that compound identification through matching observed mass spectra to libraries of known compounds is needed.

In this work, the EI MS for each chromatographically separated peak was compared against those published in the NIST EI MS library. For each one, a report is generated that contains the best matches (typically the 10 best ‘hits’ listed in a ‘hitlist’) and values of three metrics (forward match, reverse match and probability) that describe the quality of the match between the unknown MS and the library spectra. The forward (FM) and reverse matching values (RM) are derived from the modified cosine of the angle between the two spectra (normalized dot product) using all fragments from both spectra (FM) or only those present in the library spectrum (RM). Similar values for FM and RM indicates that the obtained spectra are ‘clean’, i.e., free of background interference or co-elutions. In this work, the majority (> 75 %) of the differences between the FM and RM (FM - RM) were between 0 and -20. The higher the values the better the match to the library

spectra with a value of 999 being a perfect match and a value of 0 indicating that there are no similar fragments. Typically, a value of more than 800 is used as a lower threshold to indicate that a peak has been correctly identified.²⁹ The probability is determined by comparing the differences between adjacent hits in the hitlist to get the relative probability that any hit in the hitlist is correct³⁰.

Confirming EI MS library matches and assigning molecular formulas

Figure 2 shows the EI and VUV MS for two representative peaks, labeled as compounds A and B in Figure 1a. Compound A represents a case of confirmation between the EI and VUV data while compound B represents a case where the VUV MS provides information that the EI MS cannot. Also shown are the corresponding best matching library spectra including the matching statistics (FM, RM, probability), molecular formula, compound name, observed high resolution mass, exact mass derived from the assigned molecular formula and the error in the molecular mass (in parts per million, ppm) as determined from the difference between the observed and exact masses according to the following:

$$MW_{\text{error}} = (|MW_{\text{observed}} - MW_{\text{exact}}|/MW_{\text{exact}}) * 10^6$$

Compound A was identified by EI MS library matching as 6,10,14-trimethyl-2-pentadecanone (FM = 907, RM = 910, probability = 86 %), which has a molecular formula of C₁₈H₃₆O. The observed Kovats retention index (I₁) of 1829 was in good agreement with reported values from NIST (1838 ± 17, average of 13 separate studies). The largest mass to charge (m/Q) ratio in the VUV MS was 268.272 corresponding to a molecular formula of C₁₈H₃₆O (exact mass 268.277 Da), with an error of 18 ppm. The assignment of the same molecular formula from both ionization methods provides much greater confidence in peak identification than from library matching and agreement in I₁ values alone. In total, 185 peaks were identified in this way; by matching to an EI MS library spectra and subsequent confirmation of the molecular ion by peak fitting of the high resolution VUV MS. Of these 14 were excluded because the observed I₁ values were different by more than 100 or one n-alkane carbon number compared to those reported in the NIST database (Supporting Information, Figure S1). The remaining 171 peaks are identified as EI match confirmed by VUV and are represented by the filled blue squares in Figure 1b. The corresponding VUV MS for these identified compounds are shown as an appendix within the Supporting Information and represent an extension to the published VUV MS data presented in Isaacman et al.¹⁸.

Compound B was identified as 5,7-Octadien-3-ol, 2,4,4,7-tetramethyl-, (E)- (FM = 785, RM = 793, probability = 19 %, molecular formula: C₁₂H₂₂O) from the EI MS library matching. The largest mass to charge (m/Q) ratio in the VUV MS was 168.116 corresponding to a molecular formula of C₁₀H₁₆O₂ (exact mass 168.115 Da) indicating an incorrect EI MS match. The observed I₁ of 1208 was larger than the estimated 1075 ± 39 (based on structure) from the NIST11 database⁹ also indicating the likelihood of an incorrect EI MS match. In this case, the molecular formula determined from the VUV data was assigned to this peak. The largest three fragment ions in the EI MS were identified as C₂H₃O (MW_{error} = 37 ppm), C₆H₇O (MW_{error} = 98 ppm) and C₇H₁₀O (MW_{error} = 64 ppm) (Figure 2c), consistent with the assigned molecular formula from the

high resolution VUV MS data. These provide additional information for structural elucidation, as demonstrated in other work from our group^{31, 32}, but which is beyond the current scope of this paper. It should be noted that due to the higher detection limits of the VUV ionization not all peaks observed with EI could be detected with VUV (66 were not detected under VUV ionization). In total, 264 peaks were assigned VUV formulas. Of these 53 were excluded because the molecular formula assigned from the VUV MS was determined to be incorrect because it was odd and did not contain an odd number of nitrogen atoms or that the high resolution peak fitting did not yield one unique formula or the error in the observed and exact masses were greater than 100 ppm. A further 10 were excluded because the carbon number of the identified molecular ion was not in agreement with the observed trend for species with the same KMD implying complete fragmentation of the molecular ion (see example in Figure S2, Supporting Information). The remaining 211 are identified as filled red squares in Figure 1b.

Quantitative assessment of EI mass spectral library matches

There is a possibility of an incorrect identification from library matching alone even when considering the comparability of retention indices because of the high degree of fragmentation resulting from EI and with the knowledge that the range of organic compounds present in nature far exceeds the number of compounds represented in published mass spectral libraries. Peak identification using EI MS library matching is subjective and no quantitative guidelines currently exist to assist end users. NIST provides qualitative guidelines on how to interpret the reported matching statistics (FM, RM) generated by their MS search program when comparing unknown EI spectra to EI library spectra. In general, these guidelines indicate that a value of > 900 is considered an excellent match; 800–900, a good match; 700–800, a fair match and <600 a very poor match.⁹ However, since current EI libraries do not contain spectra corresponding to all compounds there is a possibility of obtaining false positive identifications. Following these guidelines and only including those compounds that are classified as having good matches (FM, RM \geq 800), as is very often done, limits the number of peaks included to less than half of those observed (238 out of 540). Approximately one quarter (42) were incorrectly identified (Figure S3, Supporting Information) similar to the example of compound B. This has implications for substantially misrepresenting the composition of a sample. This highlights the critical importance of using a complementary technique, such as the one presented here, to confirm compound identifications, as EI MS library matching alone cannot provide unambiguous compound identification even when constrained with retention time information. Using a combination of both EI and VUV provides more than twice as much molecular level speciation information (Figure S3, Supporting Information).

Figure 3 shows the distribution of MS matches (FM) for peaks characterized as either EI match confirmed by VUV or VUV formula assigned and the likelihood that the identified match is incorrectly assigned based on confirmation of the assigned molecular formula by the molecular ion identified in the VUV spectra. This figure shows the probability of a false positive match or incorrect identification as a function of the match value: > 900 the probability of an incorrect assignment is 14 % (this may be lower for matches above 950 but there was not enough data from this work to confirm); 800 – 900, 30 %; 700 – 800,

71 % and < 600, 100 %. This work represents the first quantitative assessment of the likelihood of incorrect compound identification as a function of matching statistics. This provides end users with an appreciation for what level of confidence they should have for compounds identified through matching to EI MS libraries. The accuracy of the representation presented here should improve over time as more data is collected.

Identification of homologous series using the Kendrick Mass Defect

Plotting the Kendrick Mass Defect (KMD) against the Kendrick Mass (Supporting Information, Figure S4) provides a method to characterize the observed compounds into groups or families of generic formulae based on those identified from high resolution peak fitting of the molecular ion from the VUV MS data, which is shown in Table S1 (Supporting Information). Table S1 also shows the names of the identified compounds from EI library matching associated with each of the groups. These are useful to both understand what types of compounds are present in each family but also to assign functional groups to the identified VUV formulas. This approach also facilitates the exclusion of compounds whose molecular ion carbon number disagrees with the majority of those in the same homologous series (same KMD) and minimizes incorrect molecular ion identifications due to compound fragmentation during ionization providing increased confidence in VUV formula assignment and compound identification (Figure S5, Supporting Information).

This approach provides an improved understanding of the composition of the sample with implications for understanding the chemical formation and fate without knowing the exact compounds present and could be useful for constraining model parameterizations. This is illustrated in Figure 4 which shows the dominant formula families (those with more than 5 homologues) present in the sample along with the number of compounds in each group. Figure 5 shows the carbon number distribution of the five most abundant formula families showing the majority of observed compounds have 10 or less carbon atoms likely representing oxidation products of monoterpenes consistent with the large emission sources of these compounds from the surrounding pine forest at the sampling location.

Development of new EI MS libraries

For those peaks that have been assigned a molecular formula from the VUV MS there is now an opportunity to use this information in conjunction with GC×GC retention times and the observed EI MS to generate new entries to an existing EI MS library or an entirely independent custom EI MS library based on actual measurements of organics from a specific set of samples, in this case atmospheric aerosols. This library can be used in conjunction with other publically available libraries and applied to future samples or could be used on its own and applied to samples that are of a similar type, e.g., atmospheric aerosols. Providing a targeted library search approach would potentially minimize false positives by limiting the compounds available to match to those that have previously been observed in similar sample types as well as facilitating the identification of compounds not currently in the NIST MS and/or other MS databases. Due to the substantial number of these classified peaks these represent a substantial fraction of information that would typically be neglected.

The EI MS library determined from this work (UCBvuv_2017_v1) is intended to be in an 'open access' format such that it is available to download (https://nature.berkeley.edu/ahg/MSLibrary/UCB_VUV_Lib_v1), utilize and improved or expanded by the wider user community. The metadata includes the following: chromatographic columns used, I₁, R₂, contact details for the generator of the MS entry, conditions under which the compound was observed, VUV MW, error in high resolution peak fitting (ppm), abundance of MW mass as a fraction of total signal under VUV conditions. This library was generated by exporting each of the EI mass spectra to the NIST MS search program including associated metadata and then saving all these MS as a custom library using the librarian tool. Instructions on how to use the library with the NIST MS search program are included in the Supporting Information.

Future Applications

By combining molecular formula information inferred from high resolution peak fitting of VUV mass spectral data with both gas chromatography retention information and EI mass fragmentation patterns it is possible to characterize previously unidentified compounds in chemically relevant functional group classes (Figure 4) that could be useful for constraining simplified model parameterizations, for example in the context of the average carbon oxidation space.³³ An example of this is shown in Figure S6 (Supporting Information) which illustrates the richer dataset available when utilizing the EI+VUV approach describe in this work compared to only EI. The structural identity of observed peaks can be determined through peak fitting of the high resolution EI data providing formulas of fragments that can then be utilized to understand the fragmentation processes yielding molecular structures.^{31, 32} This is a very time consuming exercise unless automated but is very useful. Developing software tools to automate this analysis to enable its routine use is an active area of future research. The method described in this work expands the utility of molecular tracer-based approaches by allowing identification of new tracers for which full molecular identity is not necessarily known.

CONCLUSIONS

Here we have described a new approach for the identification of organic compounds utilizing a combination of GC×GC with traditional 70eV EI MS and soft ionization using low energy (10.5 eV) VUV photons. We demonstrated our methodology by applying it to identify hundreds of individual compounds in aerosol filter samples collected at Blodgett Forest, a rural site in the Sierra Nevada Mountains. Using the published NIST EI MS library and molecular formulas determined from the high resolution VUV analysis we tested and improved the confidence in the assigned EI library matches through confirmation of the matched molecular formula as observed in the VUV MS. For those compounds whose EI match was not confirmed by the VUV MS data, molecular formulas were assigned from the observed molecular ions in the VUV MS providing more than double the information on chemical speciation and increased confidence on the organic elemental composition than is possible with matching to existing EI libraries alone. In conjunction with GC retention time information the assigned molecular formulas and the major EI fragmentation ions were used to generate novel EI MS library entries. These spectra are provided here in the form of a new custom EI MS library (see

Supporting Information) that anyone is free to download and use. This 'open access' format facilitates the potential widespread use, and permits other researchers in the community to add new entries, greatly expanding the utility of this EI MS library as a community resource. The EI MS library reported here is deliberately 'dynamic' and can be continually updated by the user community as new information about the various compound entries becomes available. As a result, important and potentially unique and novel 'tracer' compounds can be identified and characterized with implications for improving our understanding of sources and transformation processes of organic compounds in the environment.

ASSOCIATED CONTENT

Supporting information.

EI MS library (UCBvuv_2017_v1). A searchable EI mass spectral library specifically generated with atmospherically relevant compounds of interest in a format compatible with the NIST MSSEARCH program is available from:

https://nature.berkeley.edu/ahg/MSLibrary/UCB_VUV_Lib_v1.

ACKNOWLEDGEMENTS

Funding was provided by the National Science Foundation (Grant #0922562). MD was partially supported by a Charles H. Ramsden Endowed Scholarship Fund summer research fellowship (Earth and Planetary Sciences, University of California, Berkeley). GIVW was supported by a NSF Graduate Research Fellowship (#DGE 1106400). The Advanced Light Source and KRW were supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

1. Goldstein, A. H.; Galbally, I. E. Known and unexplored organic constituents in the earth's atmosphere. *Environmental Science & Technology* **2007**, *41*, (5), 1514-1521. doi:10.1021/es072476p.
2. Guenther, A.; Hewitt, C. N.; Erickson, D.; Fall, R.; Geron, C.; Graedel, T.; Harley, P.; Klinger, L.; Lerdau, M.; McKay, W. A.; Pierce, T.; Scholes, B.; Steinbrecher, R.; Tallamraju, R.; Taylor, J.; Zimmerman, P. A global model of natural volatile organic compound emissions *Journal of Geophysical Research-Atmospheres* **1995**, *100*, (D5), 8873-8892. doi:10.1029/94jd02950.
3. Isaacman, G.; Worton, D. R.; Kreisberg, N. M.; Hennigan, C. J.; Teng, A. P.; Hering, S. V.; Robinson, A. L.; Donahue, N. M.; Goldstein, A. H. Understanding evolution of product composition and volatility distribution through in-situ GC x GC analysis: a case study of longifolene ozonolysis. *Atmospheric Chemistry and Physics* **2011**, *11*, (11), 5335-5346. doi:10.5194/acp-11-5335-2011.
4. Aepli, C.; Carmichael, C. A.; Nelson, R. K.; Lemkau, K. L.; Graham, W. M.; Redmond, M. C.; Valentine, D. L.; Reddy, C. M. Oil Weathering after the Deepwater Horizon Disaster Led to the Formation of Oxygenated Residues. *Environmental Science & Technology* **2012**, *46*, (16), 8799-8807. doi:10.1021/es3015138.

5. Simoneit, B. R. T. A review of current applications of mass spectrometry for biomarker/molecular tracer elucidations. *Mass Spectrometry Reviews* **2005**, *24*, (5), 719-765. doi:10.1002/mas.20036.
6. Megharaj, M.; Ramakrishnan, B.; Venkateswarlu, K.; Sethunathan, N.; Naidu, R. Bioremediation approaches for organic pollutants: A critical perspective. *Environment International* **2011**, *37*, (8), 1362-1375. doi:10.1016/j.envint.2011.06.003.
7. Schauer, J. J.; Rogge, W. F.; Hildemann, L. M.; Mazurek, M. A.; Cass, G. R.; Simoneit, B. R. T. Source apportionment of airborne particulate matter using organic compounds as tracers. *Atmospheric Environment* **1996**, *30*, (22), 3837-3855. doi:10.1016/1352-2310(96)00085-4.
8. Williams, B. J.; Goldstein, A. H.; Kreisberg, N. M.; Hering, S. V.; Worsnop, D. R.; Ulbrich, I. M.; Docherty, K. S.; Jimenez, J. L. Major components of atmospheric organic aerosol in southern California as determined by hourly measurements of source marker compounds. *Atmospheric Chemistry and Physics* **2010**, *10*, (23), 11577-11603. doi:10.5194/acp-10-11577-2010.
9. Stein, S. E. *National Institute and Standards and Technology (NIST) Mass Spectral Search Program, Version 2.0g*; 2011.
10. Welthagen, W.; Schnelle-Kreis, J.; Zimmermann, R. Search criteria and rules for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry analysis of airborne particulate matter. *J. Chromatogr. A* **2003**, *1019*, (1-2), 233-249.
11. Hamilton, J. F. Using Comprehensive Two-Dimensional Gas Chromatography to Study the Atmosphere. *Journal of Chromatographic Science* **2010**, *48*, (4), 274-282, 10A-11A. doi:10.1093/chromsci/48.4.274.
12. Liu, Z. Y.; Phillips, J. B. Comprehensive 2-dimensional gas chromatography using an on column thermal modulator interface. *Journal of Chromatographic Science* **1991**, *29*, (6), 227-231.
13. Goldstein, A. H.; Worton, D. R.; Williams, B. J.; Hering, S. V.; Kreisberg, N. M.; Panic, O.; Gorecki, T. Thermal desorption comprehensive two-dimensional gas chromatography for in-situ measurements of organic aerosols. *J. Chromatogr. A* **2008**, *1186*, (1-2), 340-347. doi:10.1016/j.chroma.2007.09.094.
14. Worton, D. R.; Kreisberg, N. M.; Isaacman, G.; Teng, A. P.; McNeish, C.; Gorecki, T.; Hering, S. V.; Goldstein, A. H. Thermal Desorption Comprehensive Two-Dimensional Gas Chromatography: An Improved Instrument for In-Situ Speciated Measurements of Organic Aerosols. *Aerosol Science and Technology* **2012**, *46*, (4), 380-393. doi:10.1080/02786826.2011.634452.
15. Eschner, M. S.; Welthagen, W.; Gröger, T. M.; Gonin, M.; Fuhrer, K.; Zimmermann, R. Comprehensive multidimensional separation methods by hyphenation of single-photon ionization time-of-flight mass spectrometry (SPI-TOF-MS) with GC and GC×GC. *Analytical and Bioanalytical Chemistry* **2010**, *398*, (3), 1435-1445. doi:10.1007/s00216-010-4021-0.
16. Hanley, L.; Zimmermann, R. Light and Molecular Ions: The Emergence of Vacuum UV Single-Photon Ionization in MS. *Analytical Chemistry* **2009**, *81*, (11), 4174-4182. doi:10.1021/ac8013675.
17. Isaacman, G.; Chan, A. W. H.; Nah, T.; Worton, D. R.; Ruehl, C. R.; Wilson, K. R.; Goldstein, A. H. Heterogeneous OH Oxidation of Motor Oil Particles Causes

- Selective Depletion of Branched and Less Cyclic Hydrocarbons. *Environmental Science & Technology* **2012**, *46*, (19), 10632-10640. doi:10.1021/es302768a.
18. Isaacman, G.; Wilson, K. R.; Chan, A. W. H.; Worton, D. R.; Kimmel, J. R.; Nah, T.; Hohaus, T.; Gonin, M.; Kroll, J. H.; Worsnop, D. R.; Goldstein, A. H. Improved Resolution of Hydrocarbon Structures and Constitutional Isomers in Complex Mixtures Using Gas Chromatography-Vacuum Ultraviolet-Mass Spectrometry. *Analytical Chemistry* **2012**, *84*, (5), 2335-2342. doi:10.1021/ac2030464.
19. Worton, D. R.; Zhang, H. F.; Isaacman-VanWertz, G.; Chan, A. W. H.; Wilson, K. R.; Goldstein, A. H. Comprehensive Chemical Characterization of Hydrocarbons in NIST Standard Reference Material 2779 Gulf of Mexico Crude Oil. *Environmental Science & Technology* **2015**, *49*, (22), 13130-13138. doi:10.1021/acs.est.5b03472.
20. Kendrick, E. A mass scale based on $\text{CH}_2=14.0000$ for high resolution mass spectrometry of organic compounds. *Analytical Chemistry* **1963**, *35*, (13), 2146-&. doi:10.1021/ac60206a048.
21. Worton, D. R.; Goldstein, A. H.; Farmer, D. K.; Docherty, K. S.; Jimenez, J. L.; Gilman, J. B.; Kuster, W. C.; de Gouw, J.; Williams, B. J.; Kreisberg, N. M.; Hering, S. V.; Bench, G.; McKay, M.; Kristensen, K.; Glasius, M.; Surratt, J. D.; Seinfeld, J. H. Origins and composition of fine atmospheric carbonaceous aerosol in the Sierra Nevada Mountains, California. *Atmospheric Chemistry and Physics* **2011**, *11*, (19), 10219-10241. doi:10.5194/acp-11-10219-2011.
22. Goldstein, A. H.; Hultman, N. E.; Fracheboud, J. M.; Bauer, M. R.; Panek, J. A.; Xu, M.; Qi, Y.; Guenther, A. B.; Baugh, W. Effects of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA). *Agricultural and Forest Meteorology* **2000**, *101*, (2-3), 113-129. doi:10.1016/s0168-1923(99)00168-9.
23. Dreyfus, G. B.; Schade, G. W.; Goldstein, A. H. Observational constraints on the contribution of isoprene oxidation to ozone production on the western slope of the Sierra Nevada, California. *Journal of Geophysical Research-Atmospheres* **2002**, *107*, (D19). doi:10.1029/2001jd001490.
24. Heimann, P. A.; Koike, M.; Hsu, C. W.; Blank, D.; Yang, X. M.; Suits, A. G.; Lee, Y. T.; Evans, M.; Ng, C. Y.; Flaim, C.; Padmore, H. A. Performance of the vacuum ultraviolet high-resolution and high-flux beamline for chemical dynamics studies at the Advanced Light Source. *Review of Scientific Instruments* **1997**, *68*, (5), 1945-1951. doi:10.1063/1.1148082.
25. Kovats, E. Gas-chromatographische charakterisierung organischer verbindungen .1. Retentionsindices aliphatischer haolgenide, alkohole, aldehyde und ketone. *Helvetica Chimica Acta* **1958**, *41*, (7), 1915-1932. doi:10.1002/hlca.19580410703.
26. Pankow, J. F.; Luo, W.; Melnychenko, A. N.; Barsanti, K. C.; Isabelle, L. M.; Chen, C.; Guenther, A. B.; Rosenstiel, T. N. Volatilizable Biogenic Organic Compounds (VBOCs) with two dimensional Gas Chromatography-Time of Flight Mass Spectrometry (GC × GC-TOFMS): sampling methods, VBOC complexity, and chromatographic retention data. *Atmos. Meas. Tech.* **2012**, *5*, (2), 345-361. doi:10.5194/amt-5-345-2012.
27. DeCarlo, P. F.; Kimmel, J. R.; Trimborn, A.; Northway, M. J.; Jayne, J. T.; Aiken, A. C.; Gonin, M.; Fuhrer, K.; Horvath, T.; Docherty, K. S.; Worsnop, D. R.; Jimenez, J. L. Field-deployable, high-resolution, time-of-flight aerosol mass spectrometer. *Analytical Chemistry* **2006**, *78*, (24), 8281-8289. doi:10.1021/ac061249n.

28. Stein, S. E.; Scott, D. R. Optimization and testing of mass spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, (9), 859-866. doi:10.1016/1044-0305(94)87009-8.
29. Kallio, M.; Jussila, M.; Rissanen, T.; Anttila, P.; Hartonen, K.; Reissell, A.; Vreuls, R.; Adahchour, M.; Hyotylainen, T. Comprehensive two-dimensional gas chromatography coupled to time-of-flight mass spectrometry in the identification of organic compounds in atmospheric aerosols from coniferous forest. *J. Chromatogr. A* **2006**, *1125*, (2), 234-243. doi:10.1016/j.chroma.2006.05.050.
30. Stein, S. E. Estimating probabilities of correct identification from results of mass spectral library searches. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, (4), 316-323. doi:10.1016/1044-0305(94)85022-4.
31. Ruehl, C. R.; Nah, T.; Isaacman, G.; Worton, D. R.; Chan, A. W. H.; Kolesar, K. R.; Cappa, C. D.; Goldstein, A. H.; Wilson, K. R. The Influence of Molecular Structure and Aerosol Phase on the Heterogeneous Oxidation of Normal and Branched Alkanes by OH. *J. Phys. Chem. A* **2013**, *117*, (19), 3990-4000. doi:10.1021/jp401888q.
32. Zhang, H. F.; Ruehl, C. R.; Chan, A. W. H.; Nah, T.; Worton, D. R.; Isaacman, G.; Goldstein, A. H.; Wilson, K. R. OH-Initiated Heterogeneous Oxidation of Cholestane: A Model System for Understanding the Photochemical Aging of Cyclic Alkane Aerosols. *J. Phys. Chem. A* **2013**, *117*, (47), 12449-12458. doi:10.1021/jp407994m.
33. Kroll, J. H.; Donahue, N. M.; Jimenez, J. L.; Kessler, S. H.; Canagaratna, M. R.; Wilson, K. R.; Altieri, K. E.; Mazzoleni, L. R.; Wozniak, A. S.; Bluhm, H.; Mysak, E. R.; Smith, J. D.; Kolb, C. E.; Worsnop, D. R. Carbon oxidation state as a metric for describing the chemistry of atmospheric organic aerosol. *Nature Chemistry* **2011**, *3*, (2), 133-139. doi:10.1038/nchem.948.

Figure 1. (a) GCxGC chromatogram of a representative filter sample analysis from the BEARPEX campaign plotted as a function of primary (t_{R1} , mins) and secondary retention time (t_{R2} , secs). Observed peaks are indicated (open squares). Two representative peaks, referenced in the text, are indicated as compound A and B. (b) Same GCxGC chromatogram as shown in (a) by replotted as a function of I_1 (Kovats retention indices) and R_2 (Pankow retention time ratio) with color coded categorization of observed peaks into 5 types as discussed in the text.

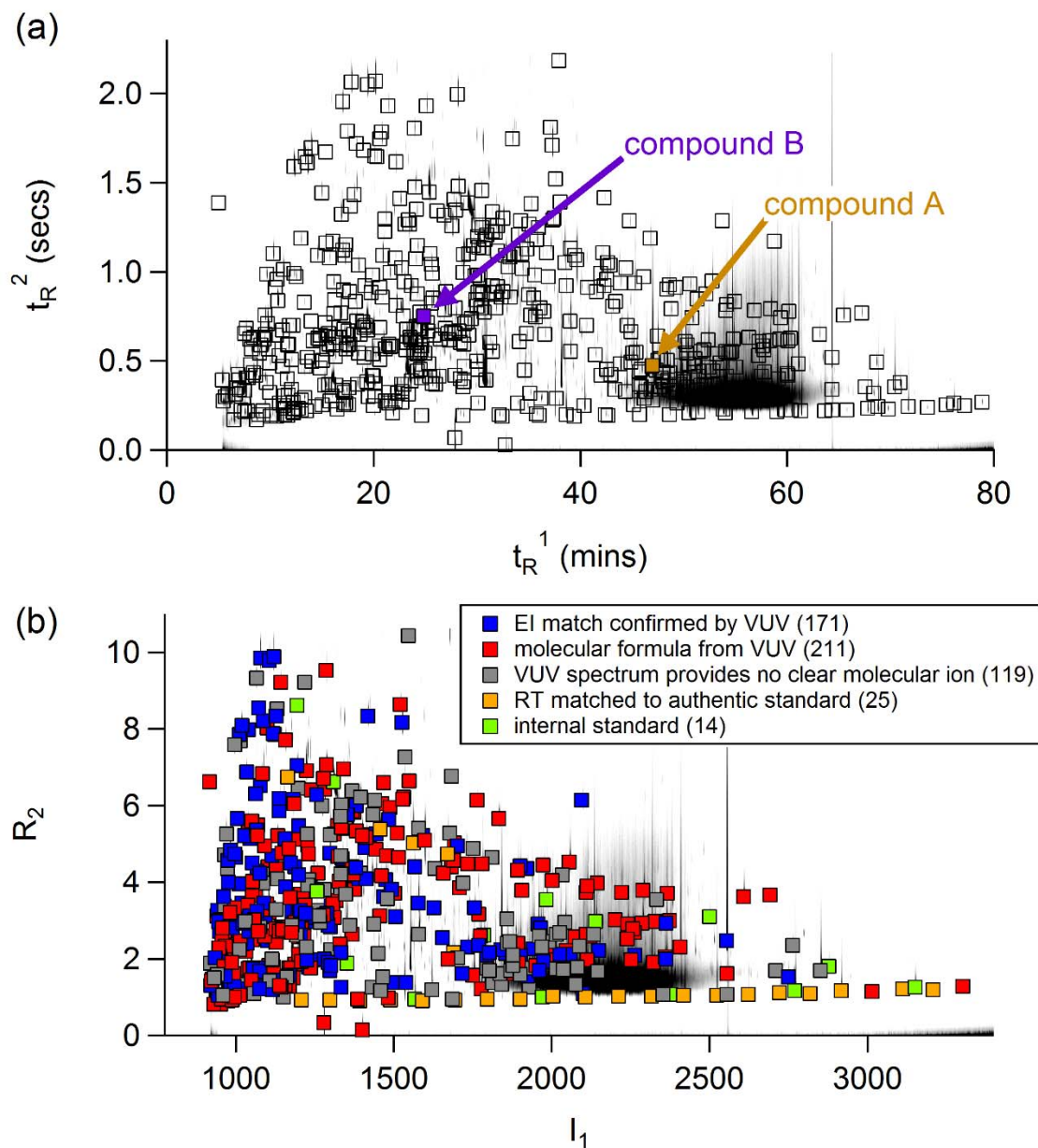


Figure 2. (a) EI MS of peak identified in Figure 1 as compound A ($t_R^1 = 46.96$ mins, $t_R^2 = 0.47$ secs) along with the library spectra from NIST identified as the best match. (b) VUV MS of compound A. A close up of the high resolution data corresponding to the molecular ion is also shown (inset). (c) EI MS of peak identified in Figure 1 as compound B ($t_R^1 = 24.85$ mins, $t_R^2 = 0.75$ secs) along with the library spectra from NIST identified as the best match. The high resolution data corresponding to the three largest fragments are also shown (inset). (d) VUV MS of compound B. A close up of the high resolution data corresponding to the molecular ion is also shown (inset).

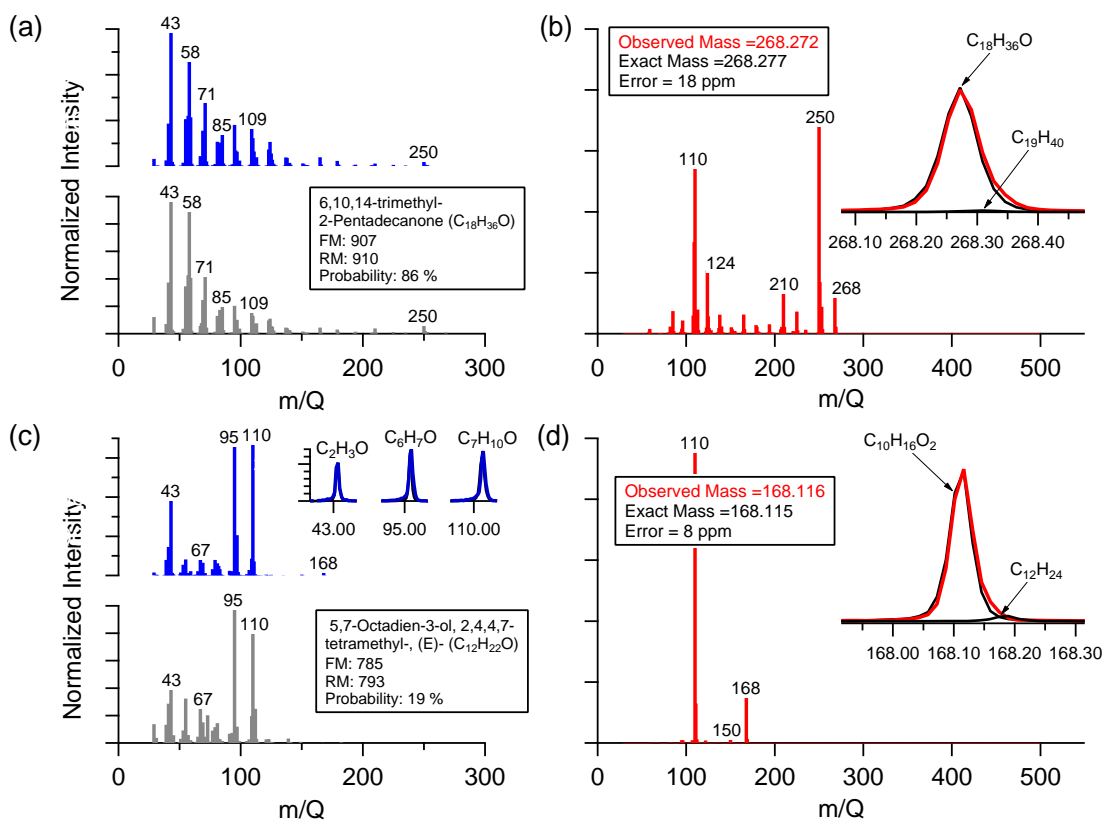


Figure 3. Distribution of observed matches to the NIST EI MS library and likelihood of correct peak assignment, based on confirmation of molecular ion by VUV mass spectra, as a function of the forward match (FM) statistic. The general guidelines supplied by NIST are also shown in red.

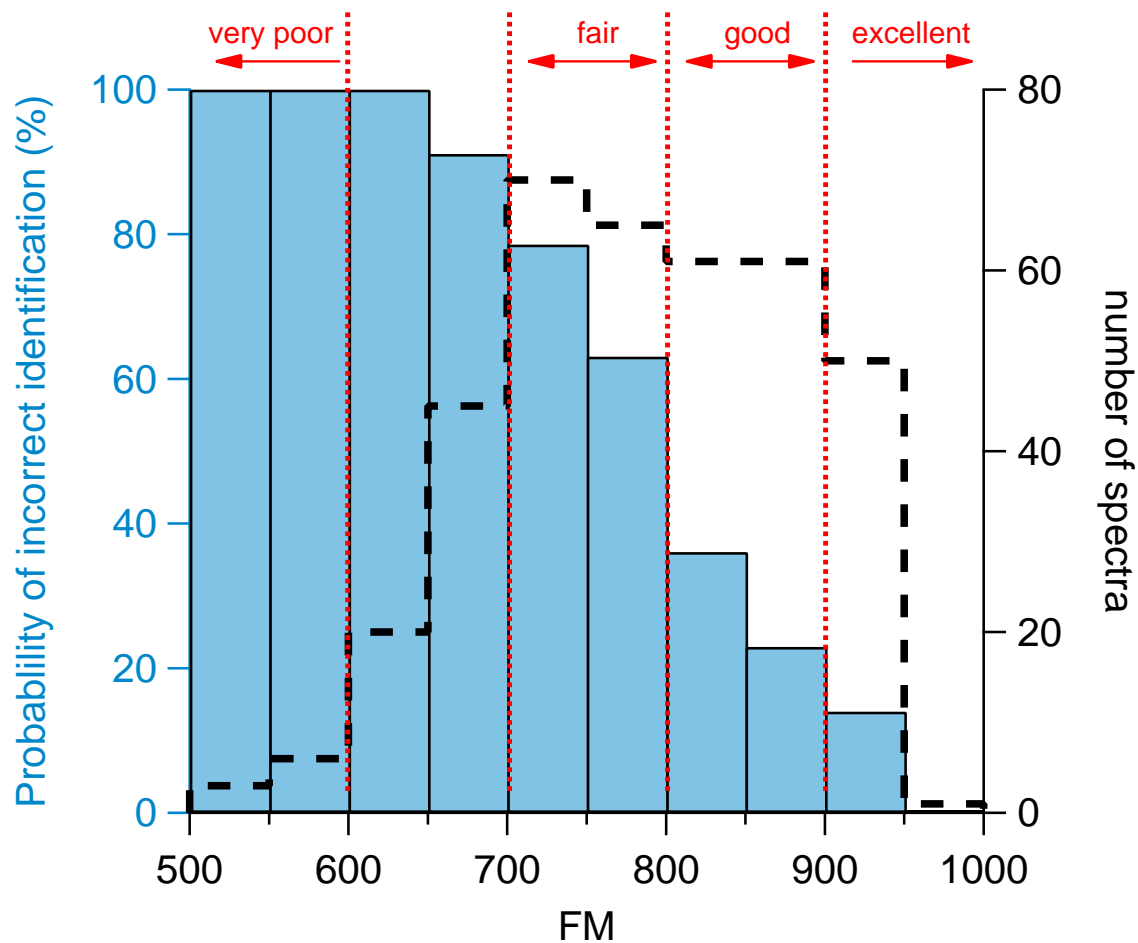


Figure 4. Number frequency distribution of molecular formulas for the major observed family classifications, those with more than 5 homologues, separated into those identified by EI MS library matching and confirmed by VUV (EI; blue) and those that were assigned formulas from the molecular ions (VUV; red).

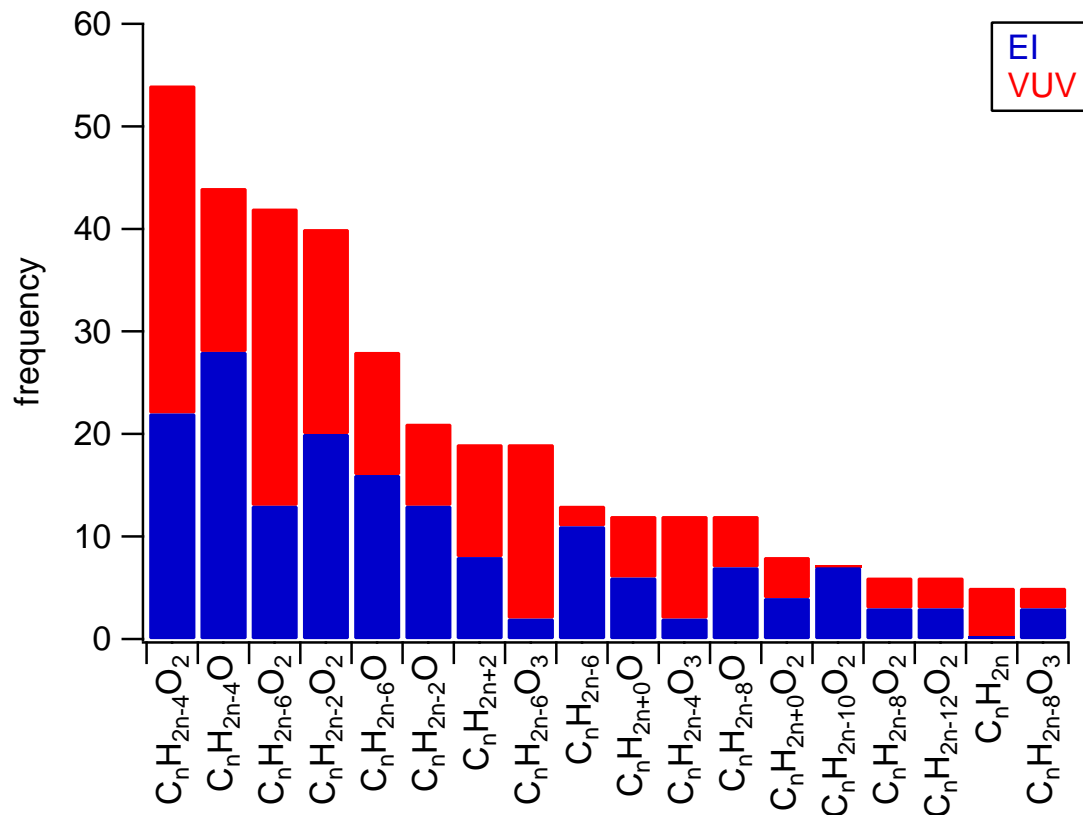


Figure 5. Number frequency distribution as a function of carbon number for the six major observed family classifications shown in Figure S2 (Supporting Information), separated into those identified by EI MS library matching and confirmed by VUV (EI; blue) and those that were assigned formulas from the molecular ions (VUV; red).

