

**An Automated System for
Validating and Enhancing GC/MS
Library Search**

Stacey Simonoff, Don Kuehl, Yongdong Wang

Cerno Bioscience, Norwalk, CT, USA

Presented at Pittcon2018

March, 2018

Orlando, FL

The Objective: Improve and automate GC/MS search and compound ID

GC/MS library search is a mainstay application for the identification, confirmation and quantitation of known and unknown compounds. In spite of its huge popularity, library search results are sometimes ambiguous and in some instances can fail to produce any useful results. This is due to a number of reasons including unresolved chromatographic peaks (mixtures), column bleed and background interferences, or simply the absence of the target compound in the library. While there are manual methods to work around some of these limitations¹, they tend to require a high degree of user expertise and can be tedious and time consuming.

Recent advances in MS calibration technology^{2,3,4} have demonstrated that commercial single quad GC/MS systems are capable of accurate mass measurements to ± 5 mDa, and more importantly the isotope lineshape profiles (Spectral Accuracy) to greater than 99.0%. This novel calibration method enables these instruments to perform reliable formula ID comparable to expensive high resolution MS instrumentation². In addition, we will introduce a new approach to peak mixture detection and spectral deconvolution which can reliably and automatically identify unresolved chromatographic peaks and deconvolve them into their pure component spectra.

In this poster, we will explore utilizing these technologies to improve library search and unknown compound ID, automate mixture and background correction, and most importantly provide orthogonal metrics to provide more accurate and robust compound ID. Finally, we will introduce a novel software platform which can fully automate the analysis.

The Solution: A multi-pronged and systematic approach

To improve the accuracy and robustness of GC/MS compound ID, a multi-pronged approach is used to mitigate the limitations mentioned previously, including:

- Identify mixture peaks using PCA and deconvolve into pure component MS
- Search “pure” MS against NIST library
- Utilize calibrated MS to perform formula ID on the molecular/fragment ion(s) as an orthogonal metric to cross validate the library search

Figure 1 outlines the algorithm⁵ used to determine mixture peaks and extract the pure spectra. In Step 1 PCA of the MS across each peak is used to calculate the number of components for each peak. Peaks that are determine to be “pure” can then be used as a model to locate the exact location and intensity of the underlying peaks in the “mixture” peaks (Step 2). Once each peak

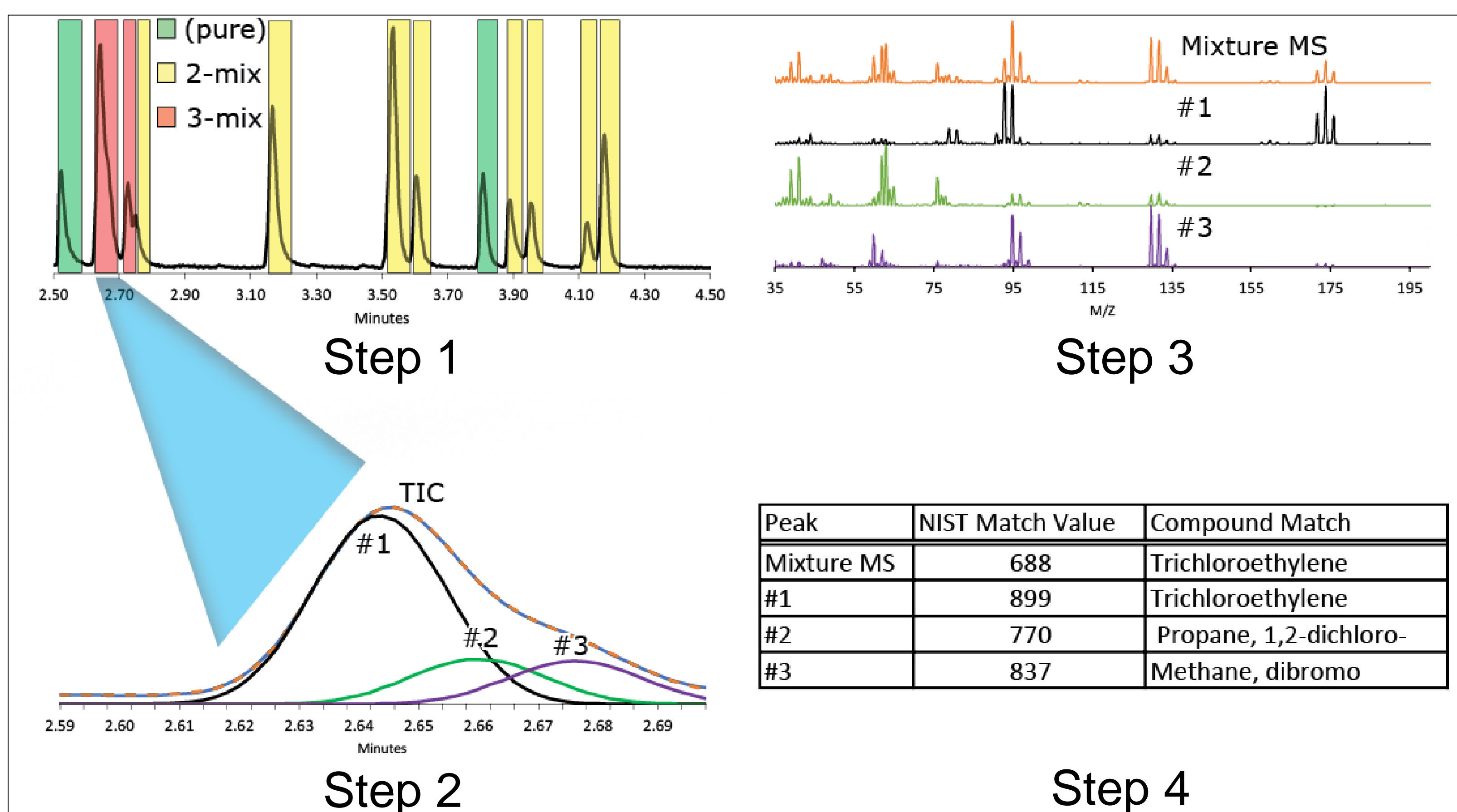


Figure 1. Steps for deconvolving mixture peaks.

position is known, a Classic Least Squares (CLS) solution can be used to obtain the “pure” spectra (Step 3). Finally, we can search the pure spectra as well as calibrate them for accurate mass (Step 4)

Next, Figure 2 illustrates the steps used to calibrate the profile mode, unit resolution data to obtain accurate mass and high Spectral Accuracy MS suitable for formula ID.

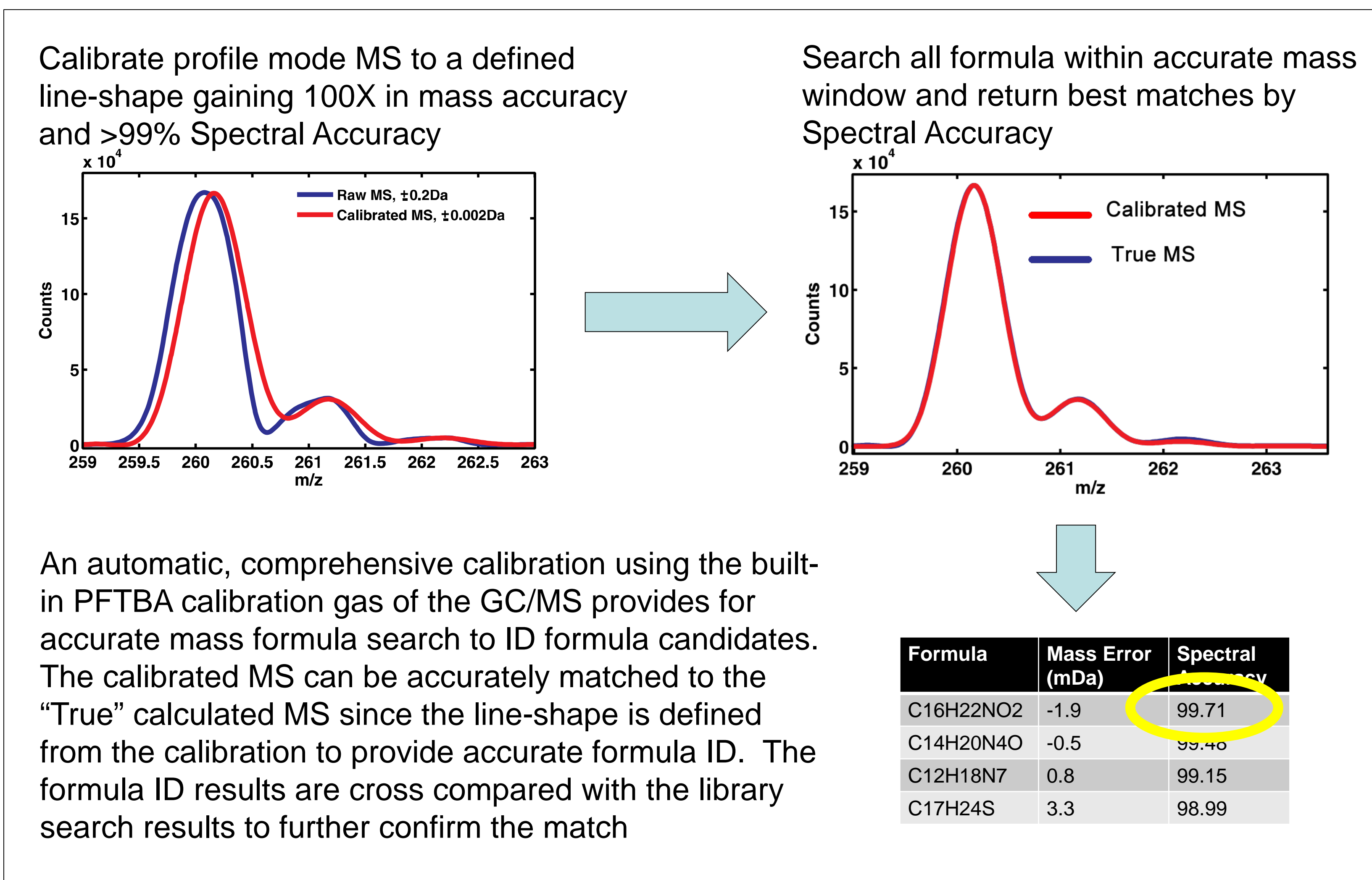


Figure 2. Steps for calibrating for accurate mass/Spectral Accuracy.

Once the spectra are properly calibrated, the formula for both molecular and fragment ions can be determined and used to further validate the compound ID. The assignment of the fragment ions is of particular importance (Figure 3), as it can assist in identifying the structure of compounds not present in the libraries.

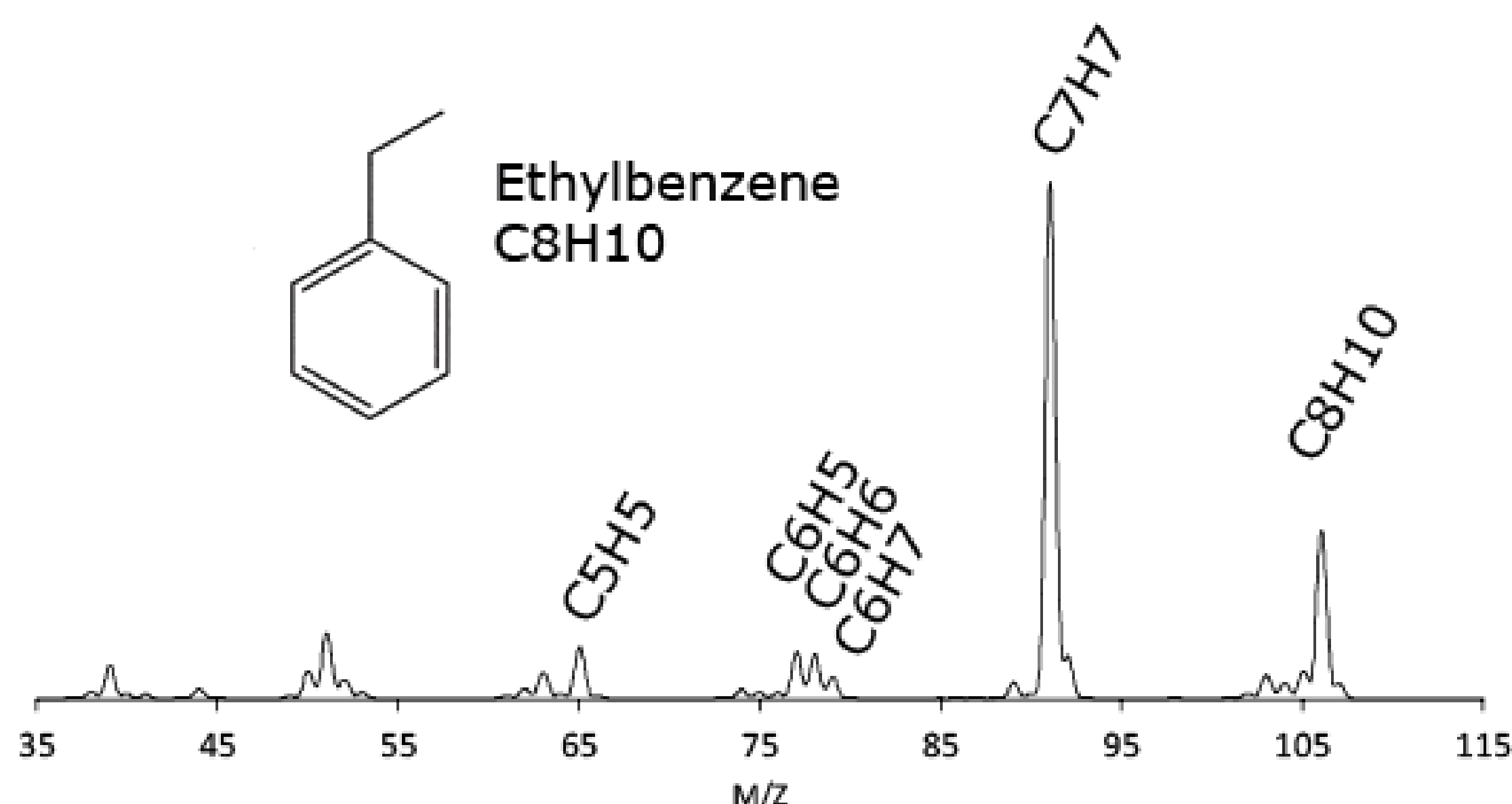


Figure 3. Formula ID of molecular and fragment ions can be powerful tools for structural elucidation.

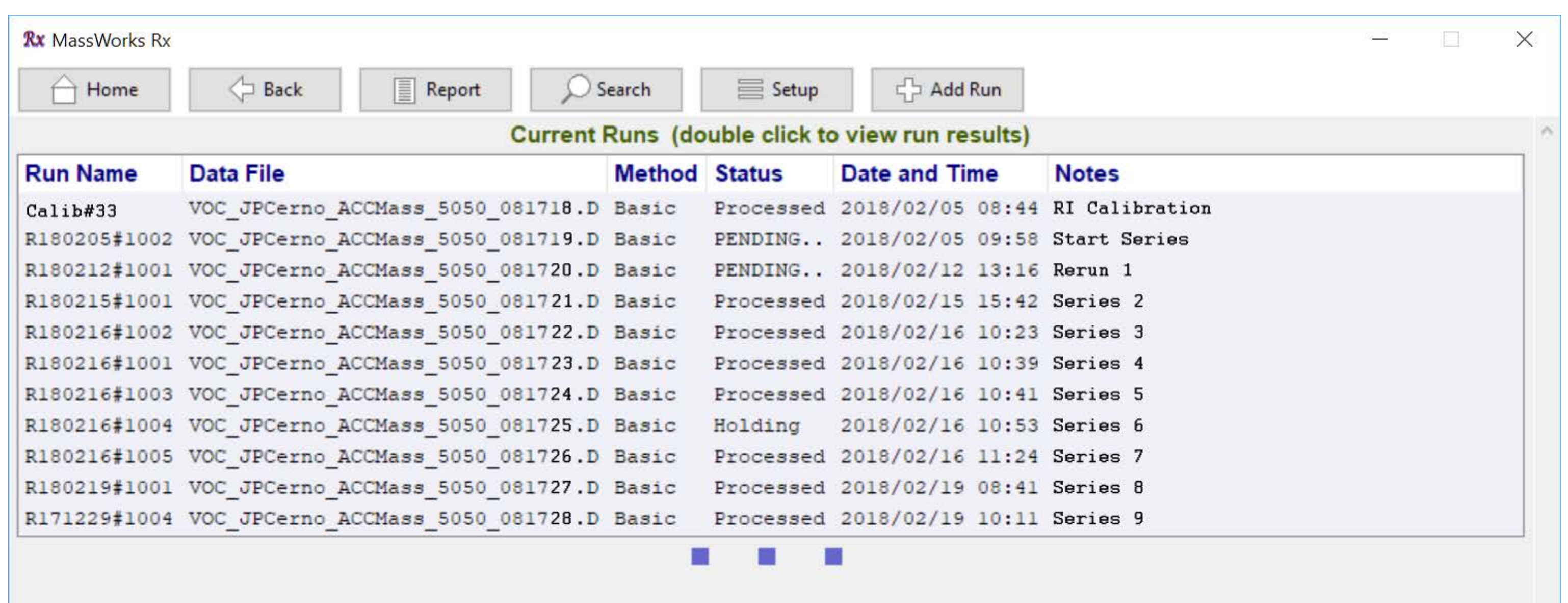
Automating the process

Without the ability to automate the described process, the technique would require manual interaction which is both time consuming and costly and thus reduces its value to many labs. To solve the problem, the algorithms were incorporated into a software architecture which is highly configurable and designed for automation, MassWorks Rx⁵. The software was configured through its built-in Method editor to perform following processing steps:

1. Data is read from the native MS vendor format
2. The TIC is used to locate peaks
3. For each peak the number of components is determined
4. The “pure” spectra are extracted from any mixture peaks
5. All pure spectra are searched against the NIST libraries
6. The MS are calibrated to accurate mass / high Spectral Accuracy
7. Formulas are determined for the molecular ion and fragments ions to verify the search results
8. Data are archived in a secure, but accessible, database

Once finished, the job is available for review and a PDF report can be distributed or archived.

The software executes the method(s) automatically as the data is available at the end of each GC/MS run. Figure 4 shows a screen capture of the method processor.



The screenshot shows the MassWorks Rx software interface. At the top, there are navigation buttons: Home, Back, Report, Search, Setup, and Add Run. Below these is a section titled "Current Runs (double click to view run results)". This section contains a table with the following columns: Run Name, Data File, Method, Status, Date and Time, and Notes. The table lists several runs, including a calibration run and a series of runs from Series 1 to Series 9. The status of the runs varies, with some being "Processed" and others "PENDING..".

Run Name	Data File	Method	Status	Date and Time	Notes
Calib#33	VOC_JPCerno_ACCMass_5050_081718.D	Basic	Processed	2018/02/05 08:44	RI Calibration
R180205#1002	VOC_JPCerno_ACCMass_5050_081719.D	Basic	PENDING..	2018/02/05 09:58	Start Series
R180212#1001	VOC_JPCerno_ACCMass_5050_081720.D	Basic	PENDING..	2018/02/12 13:16	Rerun 1
R180215#1001	VOC_JPCerno_ACCMass_5050_081721.D	Basic	Processed	2018/02/15 15:42	Series 2
R180216#1002	VOC_JPCerno_ACCMass_5050_081722.D	Basic	Processed	2018/02/16 10:23	Series 3
R180216#1001	VOC_JPCerno_ACCMass_5050_081723.D	Basic	Processed	2018/02/16 10:39	Series 4
R180216#1003	VOC_JPCerno_ACCMass_5050_081724.D	Basic	Processed	2018/02/16 10:41	Series 5
R180216#1004	VOC_JPCerno_ACCMass_5050_081725.D	Basic	Holding	2018/02/16 10:53	Series 6
R180216#1005	VOC_JPCerno_ACCMass_5050_081726.D	Basic	Processed	2018/02/16 11:24	Series 7
R180219#1001	VOC_JPCerno_ACCMass_5050_081727.D	Basic	Processed	2018/02/19 08:41	Series 8
R171229#1004	VOC_JPCerno_ACCMass_5050_081728.D	Basic	Processed	2018/02/19 10:11	Series 9

Figure 4. The Rx method processor queues up the data to be processed according to the method assigned. Data is archived in a secure, accessible database.

Summary

A collection of advanced processing methods is applied routinely and automatically to conventional single quad GC/MS data. The approach significantly enhances the confidence of the GC/MS qualitative analysis by identifying and deconvolving mixture peaks to eliminate errors in the library search. In addition, the accurate mass/Spectral Accuracy calibration provides accurate formula ID of molecular/fragment ions which also enhances confidence in the end result.

References

1. NIST MS search and NIST AMDIS for co-elution deconvolution.
2. Y. Wang, M. Gu., Anal. Chem., **2010**, 82, 7055.
3. R.J. Abel, Canadian Journal of Forensic Science, **2014**, 47, 74.
4. S. Moldoveanu & K. Kilby, ASMS Annual Meeting, 2014.
5. Patent Pending.
6. D. Kuehl, Y. Wang, S. Simonoff, Pittcon2018, Poster W415.