Improving GC/MS Library Search on a Single Quadrupole Using Complementary and Orthogonal Metrics Within the Run

Don Kuehl, Stacey Simonoff, Yongdong Wang
Cerno Bioscience, Norwalk, CT, USA

WP 313

Introduction

Quadrupole GC/MS library search has been a powerful tool for the identification of organic compounds for decades. Although additional spectral libraries have been created and hardware has been updated with better sensitivity/throughput and ease of use, no significant progress has been made to assist the analyst in determining the correct match from a long list of possible hits. This work typically requires tedious and time-consuming manual review or even re-analysis. High resolution GC/TOF and GC/Orbitrap introduce high resolution and accurate mass capabilities which enable molecular ion and fragment ion formula ID as an additional complimentary metric. However, the high cost of the instrumentation puts it over the budget of many labs. In this poster we describe a software solution (Massworks Rx GC/ID[1]) which can greatly enhance the accuracy of compound ID on a quadrupole by incorporating powerful additional metrics including accurate mass/spectral accuracy formula ID and Retention Index (RI) matching combined with a novel new algorithm for deconvolving co-eluting peaks.

Methodology

Complementary Metrics

GC/MS library search depends on the ability to accurately match the fragment "fingerprint" of "pure" EI spectra measured at a given source voltage. The search algorithms return a ranked list of most likely matches which the user must review and evaluate to verify the correct identification. In general, match quality is limited by the purity of the peak and the variability in the IE spectra, due to differences in instrumentation and measurement conditions. In addition, the nominal mass values of most commercial libraries introduce another error limiting the discriminating power of the search. Significant separation in the library match values ("breakouts") between the top match and the other matches is one indicator of a correct ID. However, the spectral "fingerprints" of isomers or similar compounds can be very similar, making the correct ID somewhat ambiguous. In some cases, knowledge of the sample chemistry can help, but there can still be significant uncertainty.

To improve library search accuracy, a logical set of complementary and orthogonal measurements and advanced processing can be applied as follows:

1. Identify mixture peaks and deconvolve to pure component spectra
2. Perform a conventional library search to produce a list of compound candidates
3. Validate each library match using accurate mass/spectral accuracy formula ID
4. Apply retention index (RI) match to provide orthogonal compound verification
5. Use accurate mass libraries to improve match quality (if available)

Calibration

While many approaches to peak deconvolution have been published, the authors have used a previously described method which uses principle component analysis to establish the number of components in each peak and then applies a novel peak fitting method to extract the pure component spectra[2]. The ability to obtain formula ID from quadrupole GC/MS has been well established[3]. This uses line-shape calibration to obtain accurate mass data combined with isotope profile matching through spectral accuracy to provide formula confirmation of the library search results. RI matching can be done in the conventional manner by running an alkanes standard. All of these steps require some form of calibration standard. The deconvolution process requires a series of "pure" peaks across a calibration run as does the RI calibration, usually a series of alkanes. Finally, the accurate mass/spectral accuracy calibration requires a series of spectral peaks across the mass range, usually PFTBA for GC/MS. These necessary calibrations can be obtained in a single run, usually refreshed once every day, and then applied to the samples being analyzed (Figure 1). This makes the calibration setup easy for the user and software can automatically process the calibration run to create the necessary calibrations.

Analysis

The entire run is automatically processed by the software. First a peak picking algorithm is used to locate peaks in the conventional manner and the mixture peaks are detected as previously described. The deconvolution step is applied to resolve the pure component spectra from any co-eluting peaks. Figure 2 illustrates the fitting process used in a typical peak containing three co-eluting species. The pure component spectra are then searched in the conventional manner using the NIST search software application programing interface. For each library match, the formula ID for the molecular ion is calculated using an accurate mass window of ~+/-20mDa and a spectral accuracy value is returned for the corresponding compound. Spectral accuracy values >98% are a strong indicator of the correct formula. If the molecular ion is insignificant probable fragment ions can also be considered.
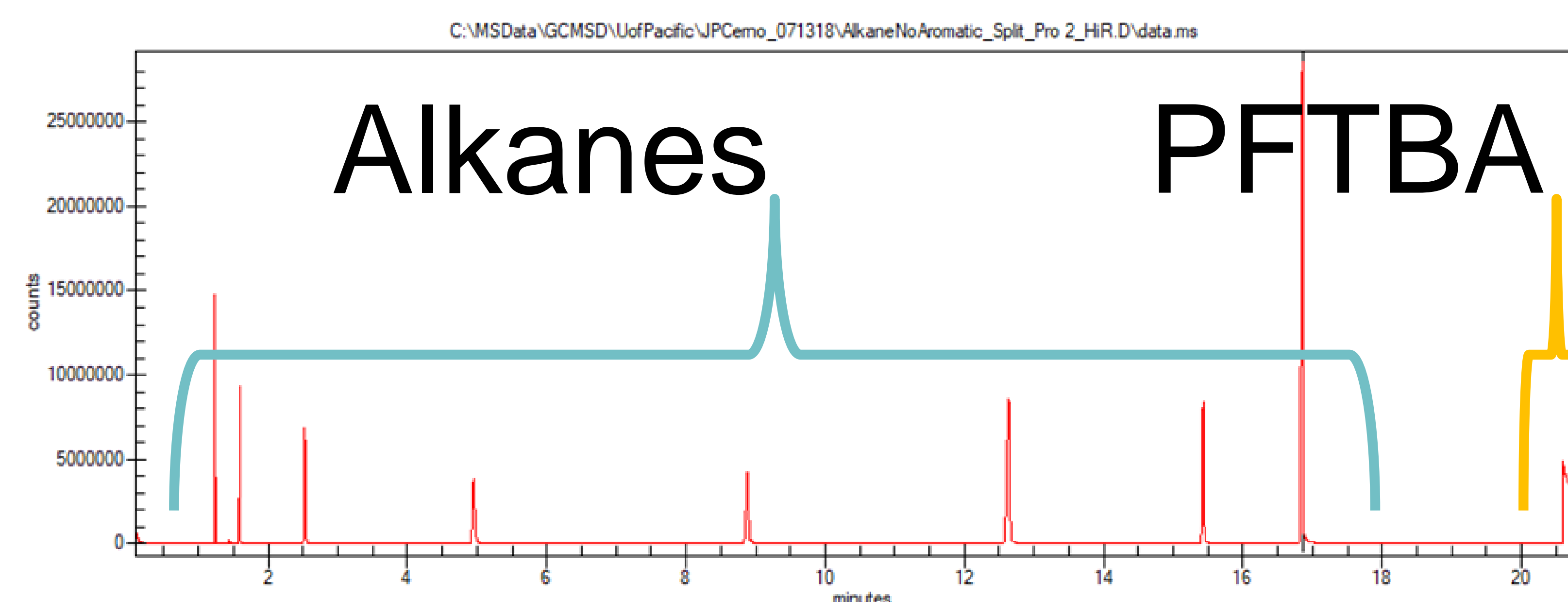

Figure 1. A single calibration run provides chromatogram calibration for peak deconvolution and RI (Alkanes) as well as a spectral calibration for accurate mass/spectral accuracy (PFTBA).
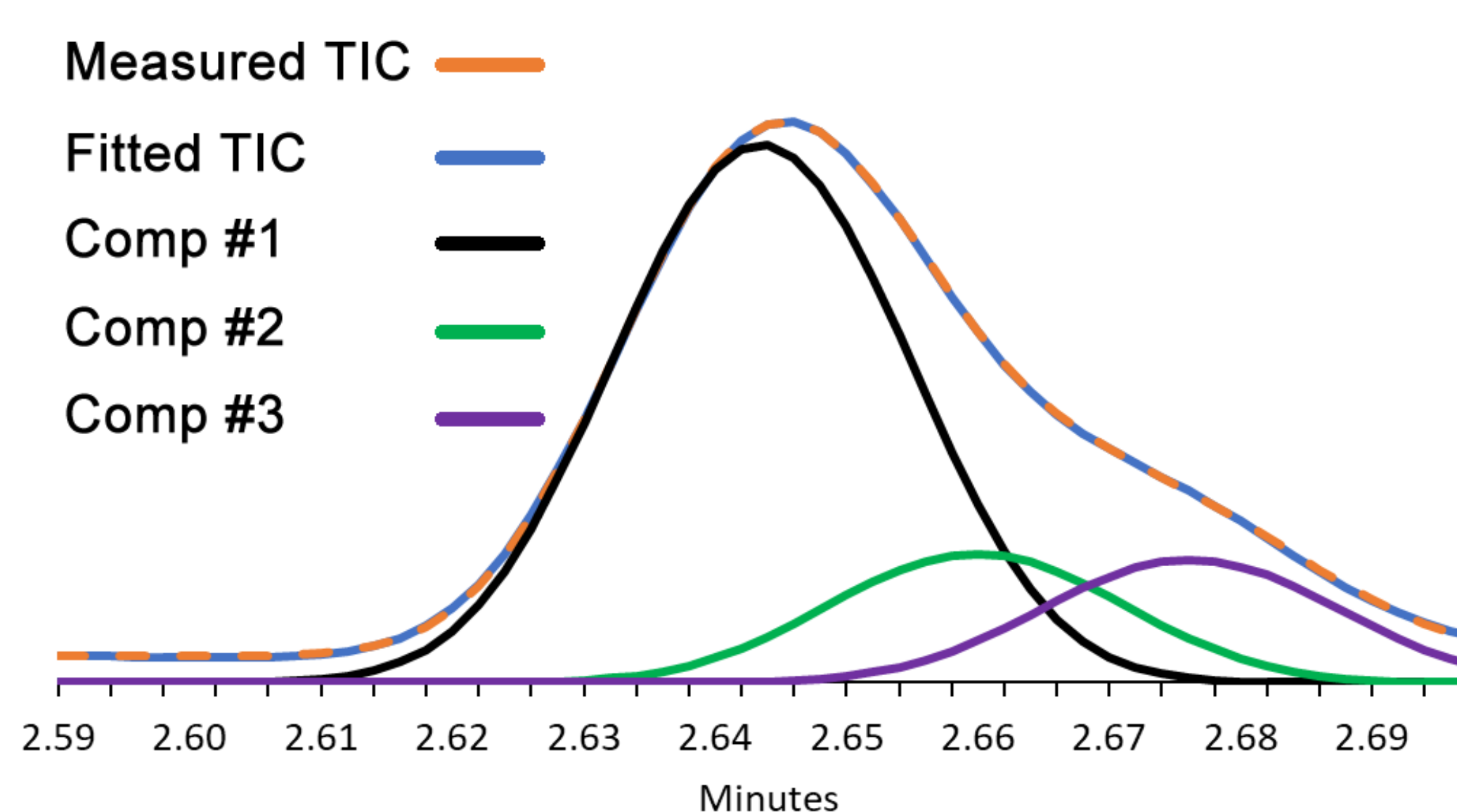

Figure 2. Plot of measured and fitted mixture peak and the underlying "pure" chromatogram peaks.

Finally, an RI value is calculated for each peak and is compared to the published value (if it exists) to generate a relative RI match valued expressed as the percentage of the shift within an alkane bracket window. It should be noted that many commercial libraries have extensive RI data, in the case of the NIST libraries, about 60% of all entries have RI data.

The software produces an interactive report which can be exported to PDF or Excel. A sample report for a VOC sample is shown in Figure 3. The top level summary shows the best match for each peak. Peaks marked in blue are "pure" and red indicates "mixture" peaks. Figure 4 dives into the full search results for peak 7 of the run which was determined to be a three component mixture. Examining the hit results shows the NIST match value (NIST) for the top compounds to be very similar. The spectral accuracy (SA) values greater then 99% confirm the correct formula for each compound, however the RI Fit value of 96.29% correctly identifies the "Z" isomer as the correct compound even though the search match value is less.
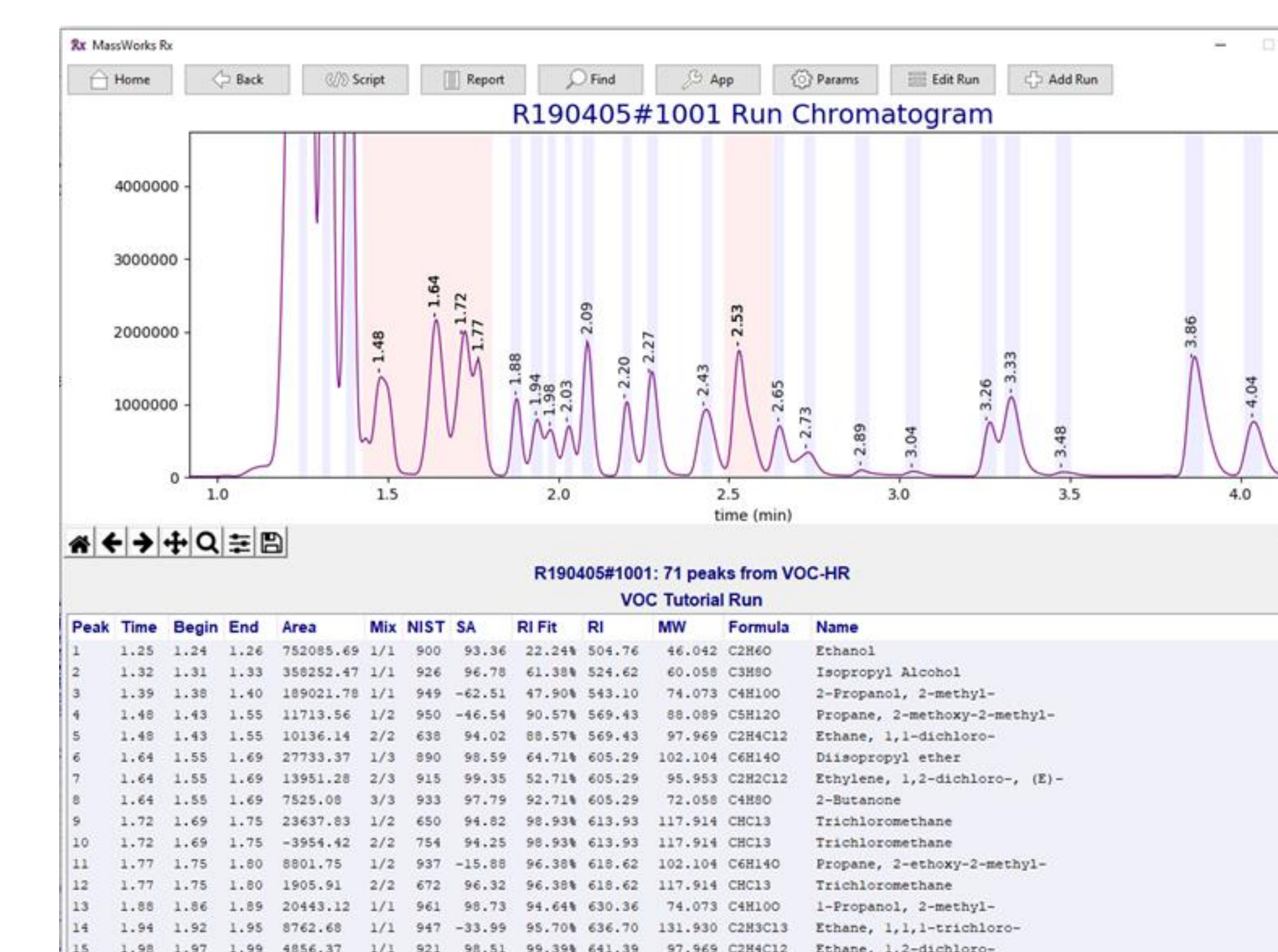

Figure 3. VOC analysis summary page showing the top matches for each peak. Red areas indicate the detection of co-eluting peaks.
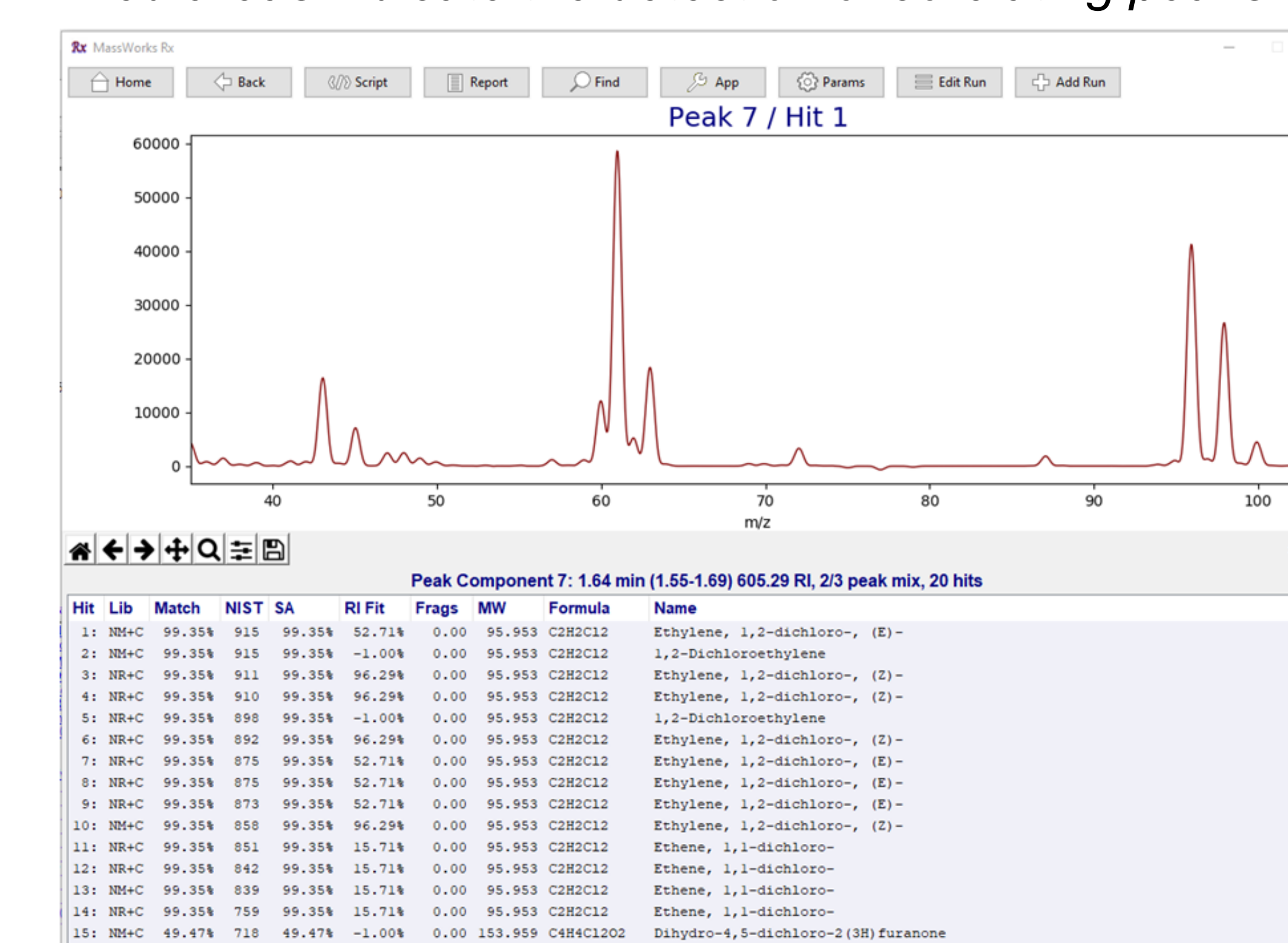

Figure 4. Search results for deconvolved mixture peak 7 showing search match (NIST), spectral accuracy (SA) and retention index match (RI Fit).

Conclusion

Providing accurate mass/spectral accuracy information for formula confirmation of quadrupole GC/MS search results provides an additional level of confirmation to compound identification. RI values give an additional orthogonal metric that further verifies the correct match with a high level of confidence, while accurate mixture deconvolution minimizes the number of unidentified peaks. Combined with automated software for processing entire runs using a single calibration, the described procedures provide an attractive method for qualitative GC/MS which can save the analyst significant time in reviewing or re-running samples without resorting to more expensive high resolution instrumentation. The accurate mass data produced by the calibration can also be used to search and/or generate high resolution libraries which can further improve the accuracy of GC/MS.

References

1. Cerno Bioscience, Norwalk, CT.
2. Y. Wang and S. Simonoff. ASMS 2018 TP816.
3. Y. Wang, M. Gu., Anal. Chem., **2010**, 82, 7055.