Automated Deformulation of LC/MS and GC/MS Data through Database Searching

Anne Marie Smith, Richard Lee, Artsiom Piatrouski, Andrey Paramonov, Vitaly Lashin
Advanced Chemistry Development, Inc. Toronto, ON, Canada

Introduction

As the capabilities of mass spectrometers (MS) progress with improved resolution and mass accuracy, the demand for analysis of samples that are more complex has emerged. Some of the sample complexity can be reduced with the use of chromatography; however, co-elution of components is nearly unavoidable.

While the use of MS instruments with greater capabilities can be used to more accurately determine elemental composition of components, these data still lack structural information. Here we describe an extension of a new chromatographic deconvolution algorithm, based on ion threads, that shows components throughout sample analysis, while also automatically searching spectral databases to aid in structure identification.

Component Detection

The feature-finding algorithm IntelliXtract 2.0 (IX 2.0) was used to generate component ion threads and mass spectra that could automatically be searched against a database (Figure 1).

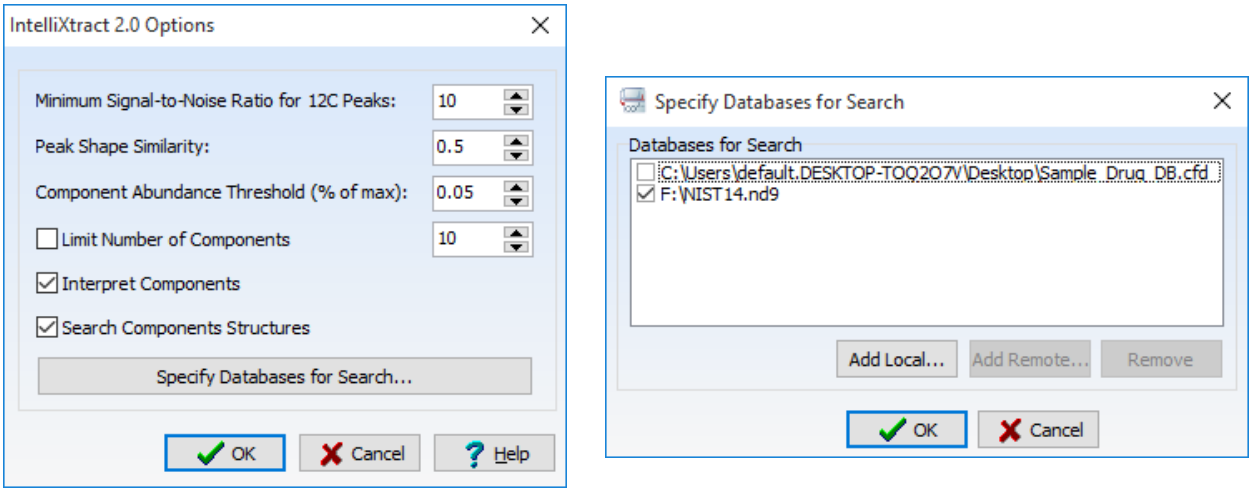

Figure 1: Settings for component detection in IX 2.0 and database search window.

Samples containing known structure-retention time pairs were analyzed to determine how well the software identified components from both liquid chromatography (LC) and gas chromatography (GC) analysis (Figure 2). LC/MS samples were searched against a database containing ESI MS2 spectra, while GC/MS samples were searched against a database containing EI MS spectra. Databases searched included a composite of commercially available databases in a *.nd9 format and internal user-created databases (*.cfd) to improve identification of components typically found within a user's common chemical space.

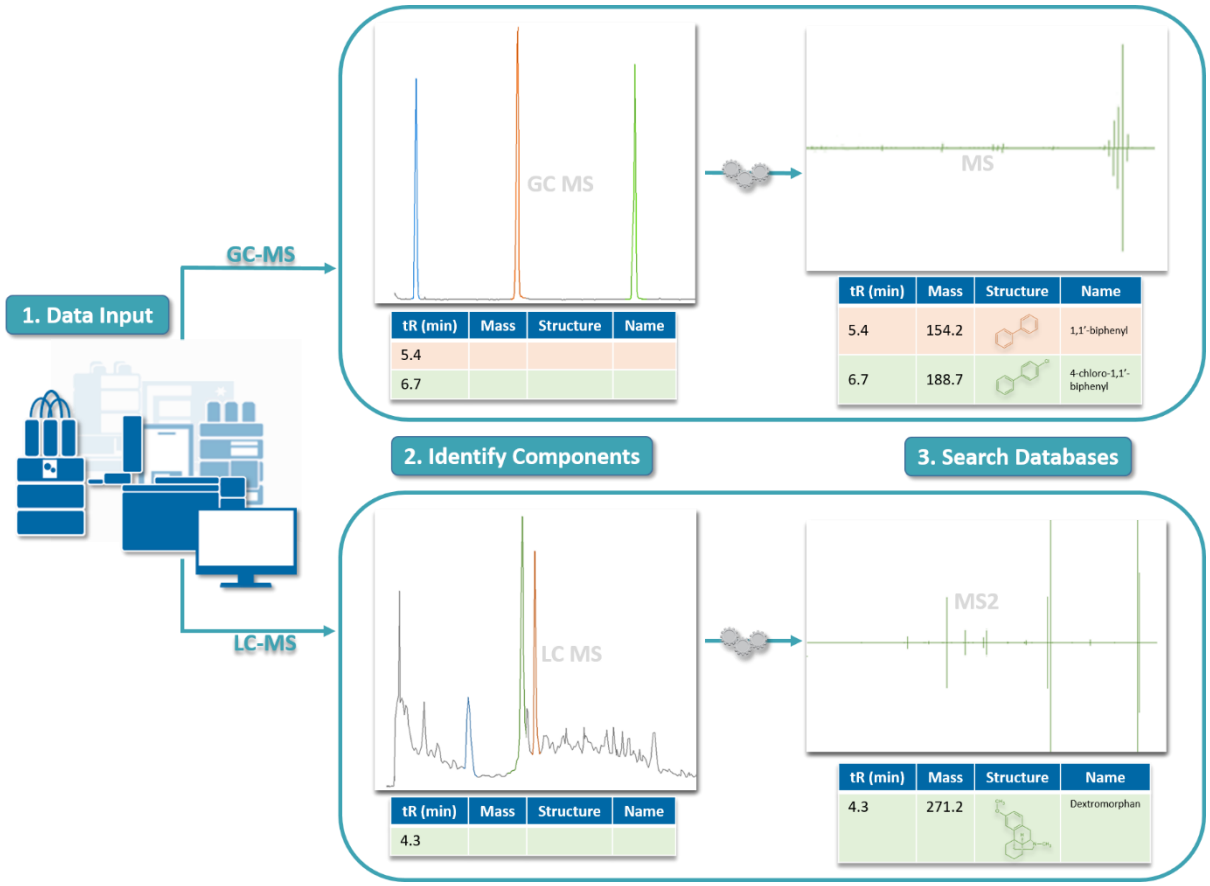

Figure 2: Model of automated deformulation workflow. Data from a GC/MS or LC/MS is opened in the software. IX 2.0 is run on either dataset to determine components present in the sample. Spectra (MS for GC/MS data or MS2 for LC/MS data) are automatically searched in specified database(s) to populate a table with top hits and provide a mirror plot of the experimental vs database spectrum.

Intelligent Component Recognition

1. Input Data

LC/MS and GC/MS data collected from a variety of vendor instruments and in various data formats can be opened in ACD/Spectrus Processor for data processing (Figure 3).

2. Identify Components

Following data import the automated feature-finding algorithm (IX 2.0) was run. The process of identifying components in the sample involved extraction of ion threads, peak integration, and grouping of spectral features to generate a component mass spectrum. Found components were populated in the Table of Components (Figure 4). The software annotated spectra and filled the table with potential fragment and adduct ions where possible.

Figure 3: Vendor data formats which can be opened in ACD/Spectrus Processor.

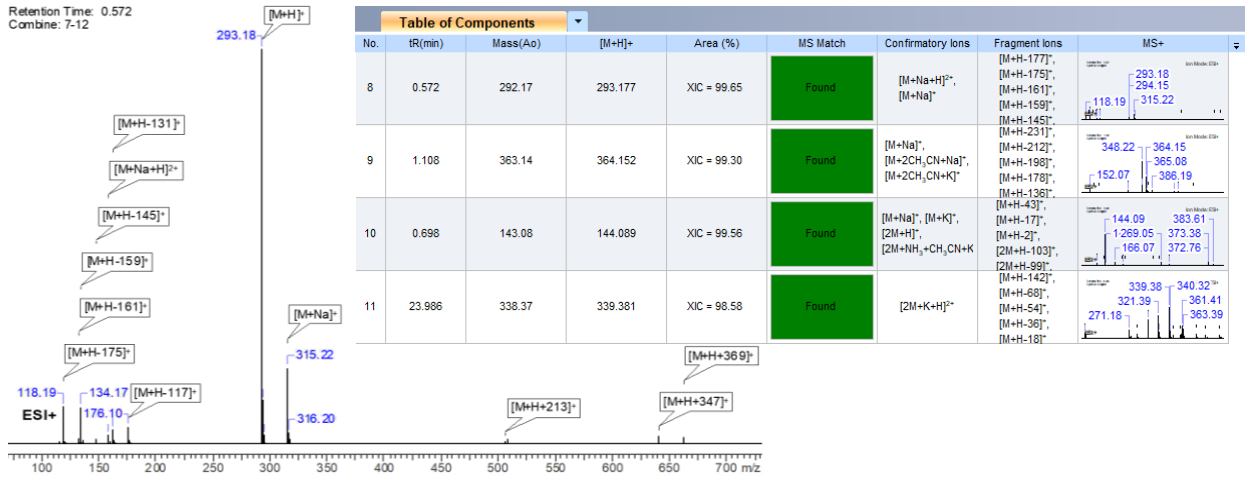

Figure 4: Table populated with components from the feature-finding algorithm prior to database searching, and a spectrum labelled with Confirmatory and Fragment Ions resulting from the sample analysis in IX 2.0. No structural information is available from the analysis at this point.

3. Search Database(s)

Prior to running IX 2.0, the user has the choice to perform a database search automatically. If the user selects to Search Component Structures, database screening would start automatically. Annotation and filling of the Table of Components would occur as before, but with additional information including the molecular formula and structure of the top database hit. The data can also be further interrogated by allowing an expert user to manually review the hit results, and select any of the returned hits. The results of a database search are displayed as a mirror plot of the experimental versus database spectra, alongside a Hit Quality Index percentage (HQI%) to allow for quick and easy assessment of match quality.

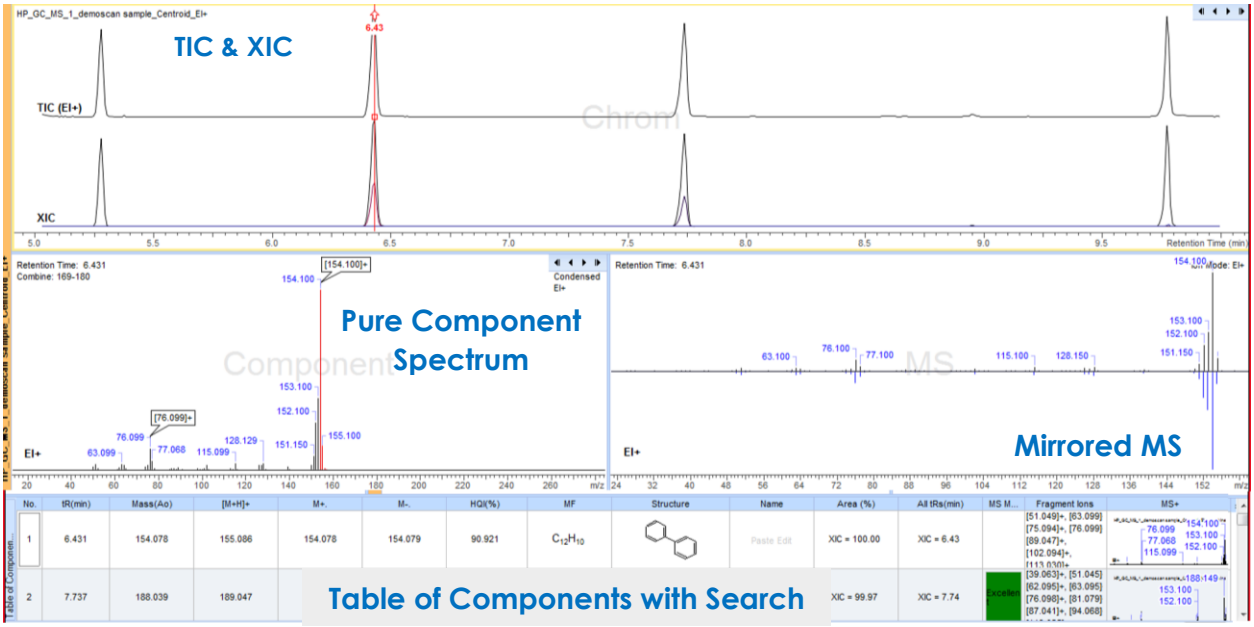

Figure 5: Screenshot of the GC/MS processed dataset with co-eluted components identified. The Table of Components shows the top hit for any given component. Selecting a specific component presents a mirrored plot for direct spectral comparison with the database hit.

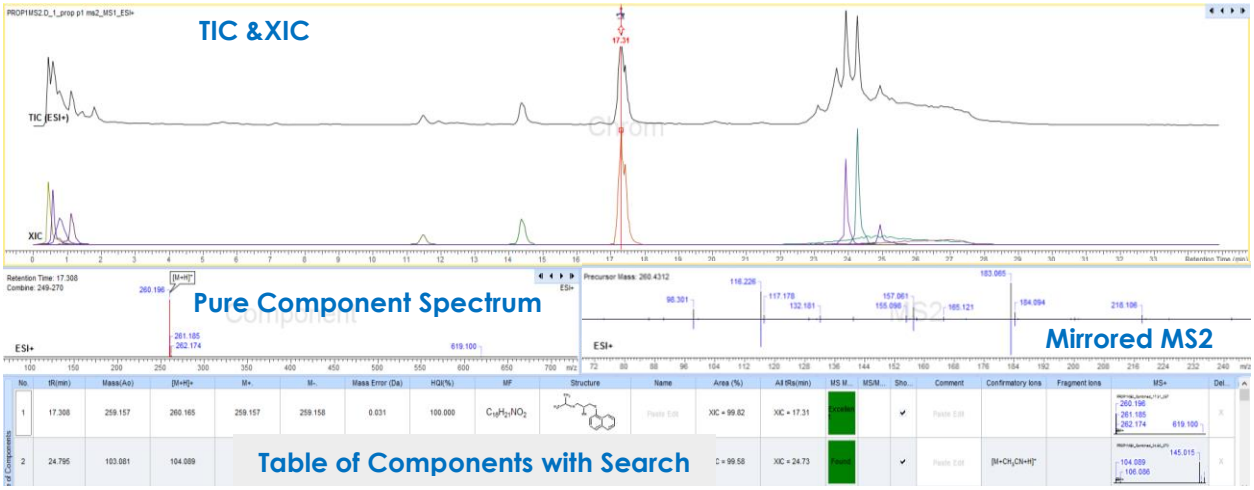

Figure 6: Screenshot of the LC/MS processed dataset. The Table of Components shows the top hit for any given component. Selecting a specific component presents a mirrored plot for direct spectral comparison with the database hit. The algorithm is able to identify co-eluting compounds and perform automated database searches, easing demands on company resources.

The resolution and size of the dataset affects the time for analysis. Deformulation results were achieved in as little as 1 minute for the GC/MS dataset shown in Figure 5, while slightly longer run times were needed for the LC/MS dataset. Users can then choose to save the processed data file for later viewing, database the results for use in future searches, or create reports.

Conclusion

The new database query feature of the deconvolution algorithm provided an important advantage in identifying structures in unknown LC/MS and GC/MS datasets — presenting an extensive, unbiased, and relevant list of structures to researchers, easing resource strain for deformulation of complex MS samples.