

ACD/LABS [ADVANCED CHEMISTRY DEVELOPMENT, INC.]

Automated Deformulation of LC/MS & GC/MS Data

Anne Marie Smith
Richard Lee
Artsiom Piatrouski
Andrey Paramonov
Vitaly Lashin

Accelerate Dereplication of MS Data

As the capabilities of mass spectrometers (MS) progress with improved resolution and mass accuracy, there is increased demand for analysis of more complex samples. While the use of MS instruments with greater capabilities can be used to determine elemental composition of components more accurately, structural information is usually still lacking.

Here, we describe a dereplication workflow using intelligent compound recognition (IXCR)—an automated workflow within ACD/MS Workbook Suite software.

An Overview of The Intelligent Component Recognition (IXCR) Workflow

The IXCR workflow can be used to help accelerate component identification for both GC/MS and LC/MS data. With a single-click you can initiate automated chromatographic feature-finding and spectral search for your entire dataset.

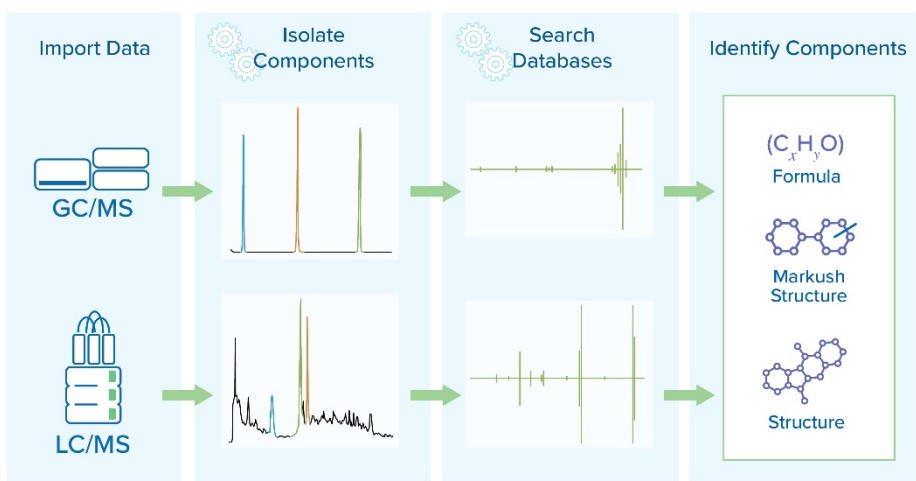


Figure 1: Import data from GC/MS or LC/MS instruments and use the automated workflow to isolate components and search databases to identify them

Applying IXCR for Component Identification

Deformation (or component recognition) can be fully automated in ACD/Labs software and executed in three steps: data import, isolation of components, and a database search that provides a list of potential candidates to accelerate component identification. These steps are discussed in more detail herein.

1. Import data from all major instrument vendors



Figure 2: MS Workbook Suite natively supports all major MS instrument vendor formats

LC/MS and GC/MS data collected from a variety of vendor instruments and in various data formats can be opened in the software for processing and analysis (Figure 2).

2. Isolate all the components in your chromatogram

The automated feature-finding IntelliXtract algorithm (IX 2.0) performs chromatographic deconvolution to isolate the components of your sample and generate a mass spectrum for each component—MS1 and MS2 when available.

Spectra are automatically annotated with potential fragment and adduct ions when possible (Figure 3).

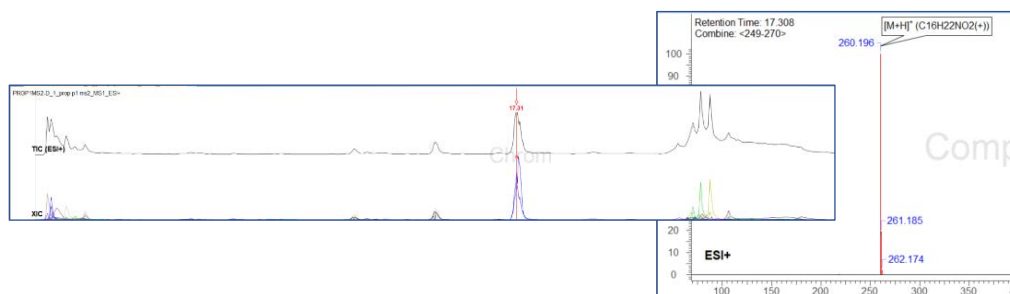


Figure 3: The deconvolution algorithm identifies co-eluting compounds (left). The mass spectrum is annotated with adduct ions (right). (Results seen after DB search)

3. Search database(s) to identify components

Both the Wiley/NIST library and user-created mass spectral databases may be used, in combination, to help with component identification.

Querying GC/MS data is straightforward because the vast majority are EI spectra where ionization energies are consistent. There is little variation between mass spectra of the same compound across different vendor instruments. Database search results can be compared and reviewed to help identify components.

The experimental spectrum and spectra resulting from the database search are presented as a mirrored plot, making it easy to compare spectra and review hit results to select the most appropriate result (Figure 4).

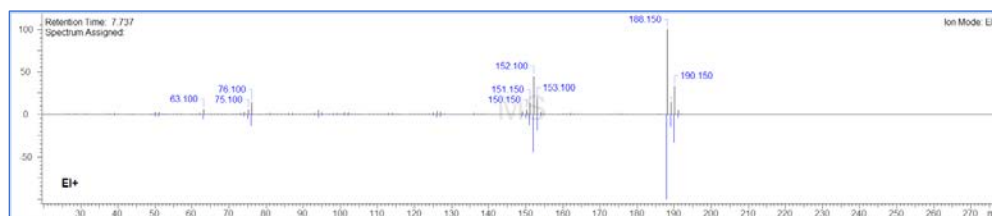


Figure 4: Search results are presented as a mirror plot to simplify review and decision-making

Instrument variability and soft ionization in LC/MS, however, requires more spectral filtering mechanisms for component identification. The software accommodates this by offering a variety of search parameters (Figure 5).

- Search MS1 or MS2 component spectra
- Search MS1 and/or MS2 reference spectra
- Search single scan or averaged spectra
- Set collision energy filters
- Set retention time tolerance filters
- Set a match-factor threshold (HQI)
- Option to apply the NIST MS Search algorithm*

*NIST MS Search can be applied by setting filters through the NIST MS Search software.

Figure 5: IXCR offers a variety of parameters to refine database search for LC/MS data

The software generates a table of components that is displayed with the chromatogram, mirror plot presentation of experimental spectrum and the database search spectrum and all other information from the database. Query results include structures when available in the database. MS Workbook Suite supports Markush structures (indicated in red on select atoms—Figure 6A) for when partial structure information is known.

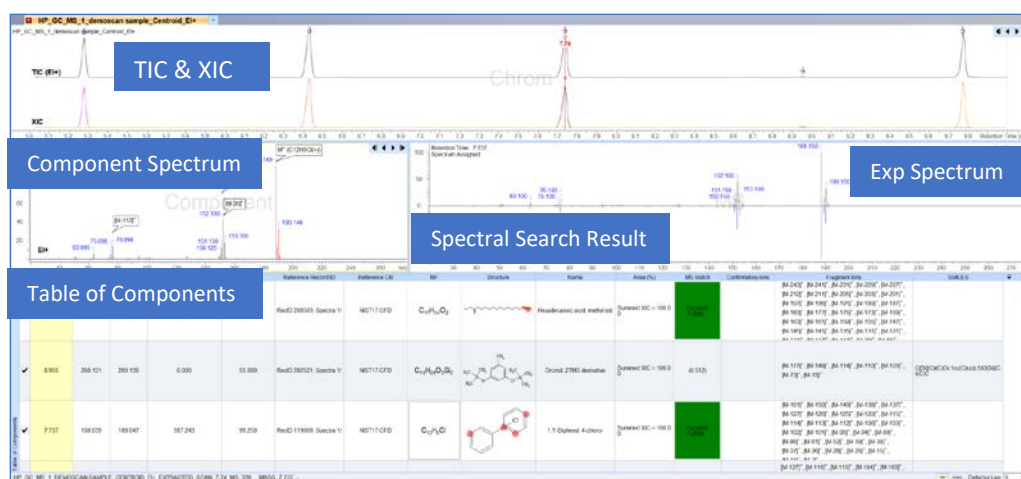


Figure 6A: Results from IXCR using the native spectral search in MS Workbook Suite include molecular formula, structure, the name of the reference database, and a measure of spectral similarity. Mirror plot presentation of experimental and the database search spectra make spectral comparison and decision-making easy.

Information present in the database is returned with search results. In the example used here the three top hits for the component spectrum were the same formula but different isomers (Figure 6B).

	IR(min)	Mass(Ao)	[M+H] ⁺	Mass Difference (ppm)	HQI(%)	Reference Record ID	Reference DB	MF	Structure	Name
✓	7.737	188.039	189.047	587.243	98.173	RecID 119087, Spectra 1/	NIST17.CFD	$C_{12}H_9Cl$		3-Chlorobiphenyl
✓	7.737	188.039	189.047	587.243	99.044	RecID 119081, Spectra 1/	NIST17.CFD	$C_{12}H_9Cl$		1,1'-Biphenyl, 2-chloro-
✓	7.737	188.039	189.047	587.243	99.259	RecID 119089, Spectra 1/	NIST17.CFD	$C_{12}H_9Cl$		1,1'-Biphenyl, 4-chloro-

Figure 6B: Expansion of the table from Figure 6A shows the top 3 returned component hits that are isomers.

Results can be reviewed manually and curated if desired. Selecting any component in the table displays the labelled component spectrum, a mirror plot of the experimental versus database spectra, and values to help you evaluate spectral match (HQI% or MF when using NIST MS Search). Depending on the complexity of the data and the database(s) used, results can be further filtered by collision energy or retention time.

The processed data file can be stored, database the results for use in future searches, or create reports.

Conclusion

The automated IXCR workflow for deconvolution of LC/MS or GC/MS datasets and database query offers efficiency and confidence in component identification for unknowns. The software presents an extensive, unbiased, and relevant list of structures or component names, easing resource strain for deformation of complex MS samples.