Fragmentation Trees for Automated de novo Interpretation of Impure Electron Ionization Spectra from Gas Chromatographic Complex Mixture Analysis—Chemical Deconvolution

Kevin Siek[1]; Vasily Makarov[2]; Viatcheslav Artaev[1]; Dmitry Mazur[3]; Albert T. Lebedev[3] | [1]*LECO Corporation, Saint Joseph, MI USA;* [2]*Mass Spectrometry Consulting Ltd., Bar, Montenegro;* [3]*Moscow State University, Moscow, Russian Federation*

Introduction

- Impure spectra contain signals from multiple analytes or signals unrelated to the main analyte, due to propinquitous coelution that confounds purely mathematical de-coelution. Some such spectra are expected in complex sample analysis.
- Impure spectra may impede manual or automated interpretation and beget false assignments or other failures.
- Chemical information within high-resolution spectra can assist mathematical de-coelution and facilitate analyte identification.

Methods

- Samples were prepared using conventional practices and analyzed by a Pegasus® GC-HRT (see handout for details).
  - *Lake Ontario fish tissue*
  - *Municipal wastewater*
  - *Nigerian crude oil*
- The wastewater sample was subsequently analyzed by *Pegasus* GC-HRT 4D to confirm results (details in handout).
- Chromatographic feature finding was performed using commercially available software (ChromaTOF-HRT® 1.81).
- Peak tables were sent to chemical deconvolution software.

Chemical Deconvolution (CD) Algorithm Steps

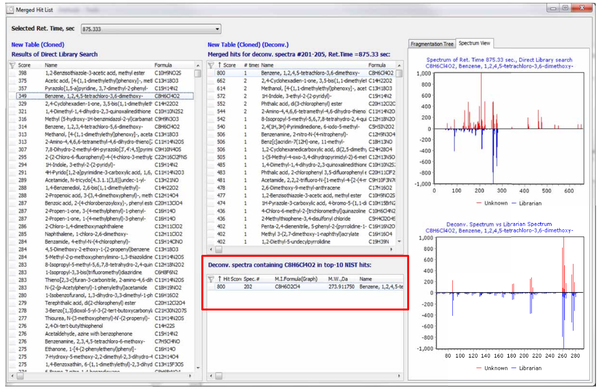
- Deisotope:
  - *Assign 2+ and 1+ charge states*
  - *Assign provisional formulas to monoisotopic signals*
- Assemble fragmentation graphs, constraining connections by:
  - *Chemically possible mass defect changes*
  - *Rational RDBE, halogen count, and heteroatom count limits*
  - *Other ad-hoc rules, such as forbid CH_2 loss*
- Analyze each fragmentation graph in the spectrum separately:
  - *Remove "orphans" (one-point, no-edge fragmentation graphs)*
  - *Find best molecular ion or top fragment formula for each graph*
  - *Library search each graph and compare to full spectrum search*
- Return results for each input spectrum:
  - *Elemental compositions of molecular ion(s) or top fragment(s)*
  - *Merged library hits, filtered by minimum elemental composition*

Results Summary

- False signals rejected from 40 hand-selected impure spectra
  - *40 spectra with total of 64 spurious signals at high m/z and intensity*
  - *55 of these 64 spurious signals were orphaned*
  - *5 of these 64 were joined to a separate minor fragmentation graph*
  - *4 of these 64 joined to the principal analyte's fragmentation graph*
- Fish tissue example coelution
  - *Incurred tetraCl-diMeO-benzene under fish fat and siloxane peaks*
  - *CD separated analyte's fragmentation graph from matrix graphs*
  - *Filtering library hits by formula yields only the one correct hit*
- Wastewater example coelution
  - *Full spectrum reflects only benzothiazole, overlooking other analytes*
  - *CD reports 3 analytes and 1 matrix component within this peak*
  - *In GC-HRT 4D analysis, same 3 analytes separate in 2nd dimension*
- Crude oil example coelution
  - *Coelution of 5 analytes sorted correctly*
  - *CD generally finds chemically distinct analytes, as in the example*
  - *CD generally fails for chemically similar analytes, as expected*

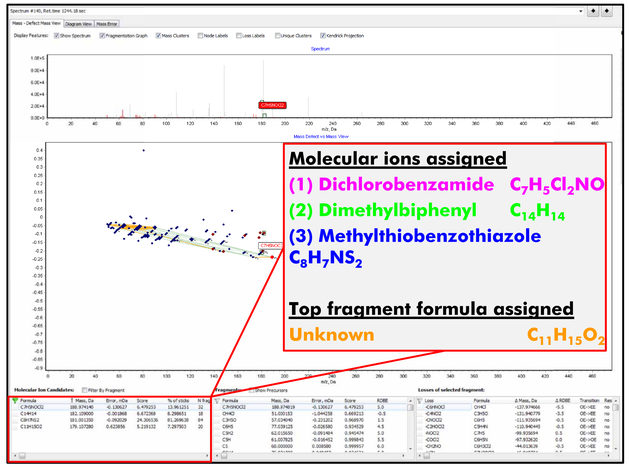
Results Example I: 3x Coelution in Fish Tissue Matrix

CD returns the one correct hit after filtering by formula; matrix (low m/z) + siloxane (high m/z) interference removed:

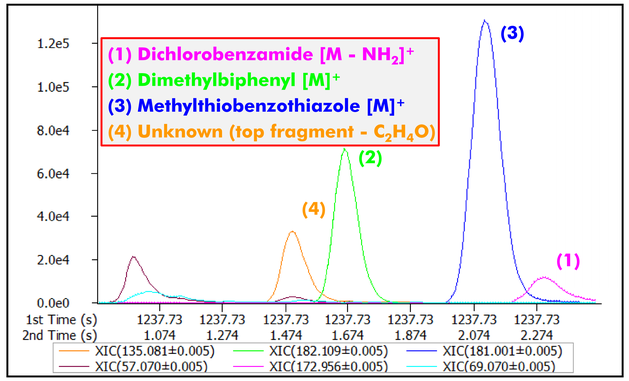


Results Example II: 4x Coelution in Wastewater

CD distinguishes 3 molecular ions and 1 top fragment:

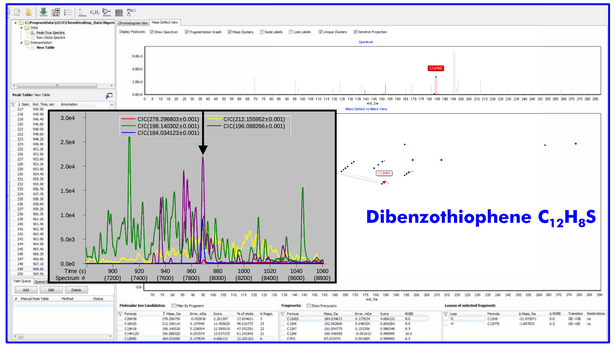


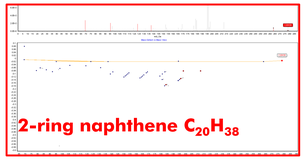
Same analytes found in subsequent GC-HRT 4D analysis:

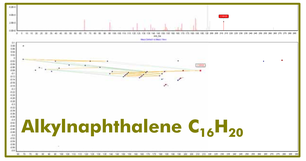


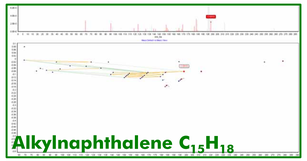
Results Example III: 5x Coelution in Crude Oil

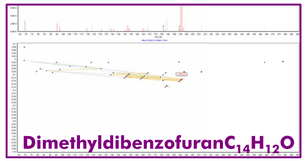
Five separate molecular ions found in coelution at ≈ 969 s:











Conclusions

- Chemical deconvolution algorithms accurately reported independent components of dead coelutions where such components belong to distinct chemical classes.
- Present algorithms fail to distinguish chemically similar analytes such as branched and linear alkanes, thus chemical deconvolution is not a substitute for GCxGC.
- Chemical deconvolution algorithms accurately rejected more than 80% of spurious signals from manually curated spectra found to be contaminated with unrelated signals.