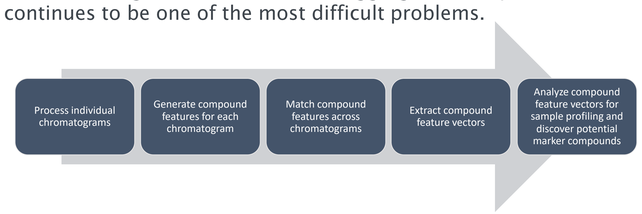
An Integrated Approach for Alignment using Peak-Region and Peak-Table Features for GC-MS and GCxGC-MS

Qingping Tao;[1] Daniel Geschwender;[1] Chase Heble;[1] Stephen E. Reichenbach;[1,2]
[1] GC Image, LLC, Lincoln NE; [2] University of Nebraska, Lincoln NE

Introduction

A key challenge for non-targeted, cross-sample analyses of gas chromatography/mass spectrometry (GC-MS or GCxGC-MS) data lies in finding the correspondence between compound features across samples. These analyses are conducted with a data analysis workflow that compares multiple samples to determine similarities and differences for sample classification and biomarker discovery. The center of this workflow is to align compound features across all chromatograms and extract an aggregated compound feature table, which continues to be one of the most difficult problems.



Compound Features

There are different types of features used for non-targeted cross-sample analyses with GC-MS **[1]** and GCxGC-MS **[2]**. Two most common types are:

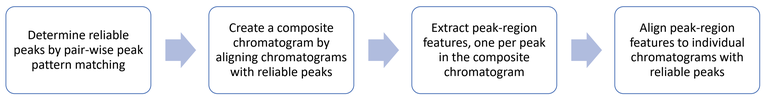
Region-based Features: Region features encompass separate retention regions of a chromatogram. However, inaccuracies may result from targeted peaks falling outside the expected region or multiple peaks eluting in a region.

Peak-based Features: Peak features capture data from peaks detected in individual chromatograms (and stored in peaks tables for each chromatogram). They are more specific than peak-region features, but have no implicit correspondence across samples. Analysis using peak features requires comprehensive peak matching, often an intractable problem with even a modest number of peaks and samples.

Investigator Workflow with Peak-Region Features

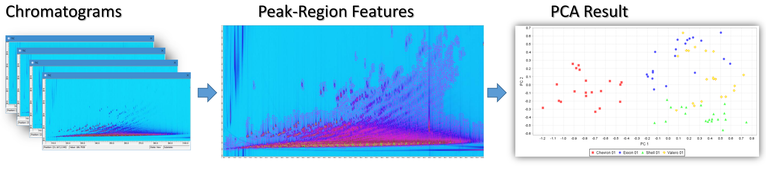
Our Investigator framework **[3,4]** combines peaks and regions into peak-region features that comprehensively captures the pattern of peaks across all sample chromatograms.

- Each peak defines a region detected in a chromatogram;
- Peak-region features are delineated by peak detection in the composite of all chromatograms;
- For each sample chromatogram, peak-region features are aligned with the detected peak pattern and generate a set of feature measurements for cross-sample analyses.



Example: Non-targeted analysis of 4 brands of diesel fuel with GCxGC-MS **[5]**.
❖ Instrument: Shimadzu 2010 Ultra GC/MS with Zoex ZX2 thermal modulator
❖ 10 samples were collected for each brand, for a total of 40 chromatograms.

Chromatograms | Peak-Region Features | PCA Result



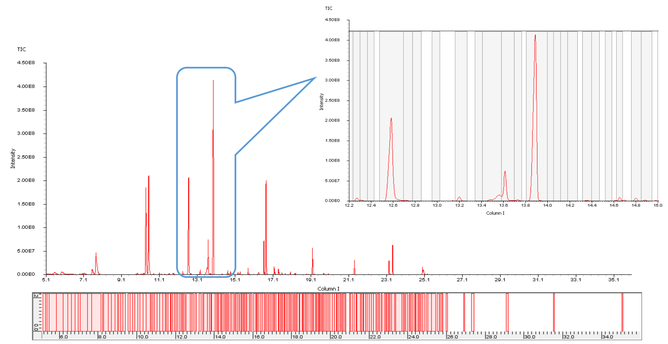
GC IMAGE
Software for Multidimensional Chromatography

University of Nebraska Lincoln

Example 1: GC-MS

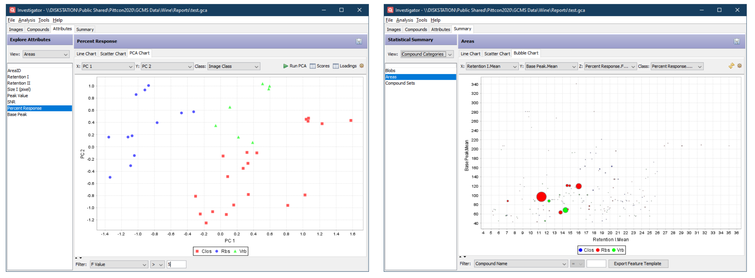
❖ Public Data Set: 39 chromatograms of wines fermented with three different commercial yeast products acquired with Agilent GC-MS system**[1]**.
❖ Data Processing: Peak detection, library search, peak alignment were all performed with an alpha version of GC Image v2.9r3.

Alignment Results:
- 176 peak-region features were extracted automatically.



➢ The extracted segmentation of the composite chromatogram

- F-value based analysis reveals potential marker compounds.



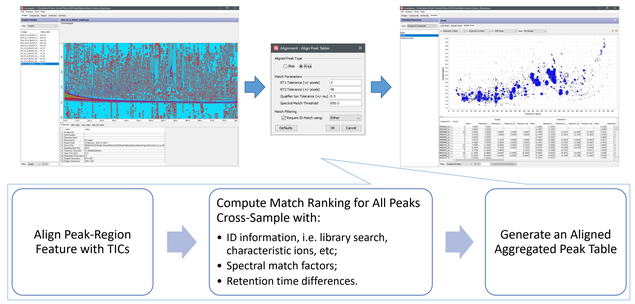
➢ Left: PCA result of all features with F value > 5.
➢ Right: A bubble plot shows all compound features with F values as bubble sizes and color assigned by the class that has larger Fisher ratios against other classes of samples.

- Verify potential marker compound features with percent response distribution and spectra from all samples.

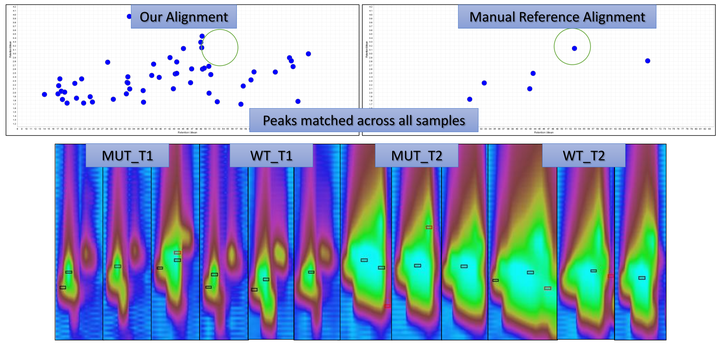
Example 2: GCxGC-MS

❖ Public Data Set: 12 chromatograms of Chlamydomonas reinhardtii alga acquired with LECO GCxGC-MS system**[6]:** two strains (MUT, WT), two time points (T1, T2), three replicates of each combination.
❖ Data Processing: Peak detection and library search were performed with vendor software. Peak alignment was performed with an alpha version of GC Image v2.9r3.

Alignment Results:
- 4319 peak-region features were extracted automatically.

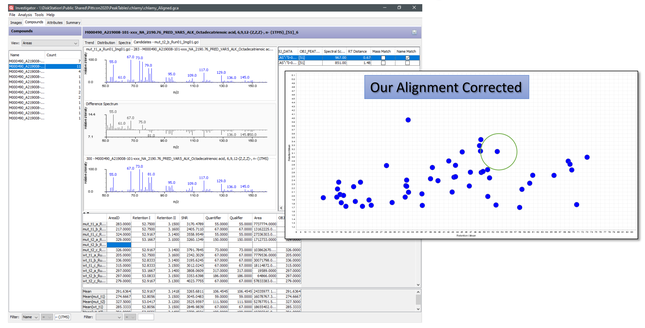


- Our alignment finds 5 of the 6 common peak groups from the manual reference; The missing peak group was split due to a split peak detection



➢ Visual confirmation of the split peaks. Black peaks are all labeled 'Octadecatrienoic acid, 6,9,12-(Z,Z,Z)-, n- (1TMS)'

- Interactive interface allows for match candidates to be manually reviewed and reassigned to achieve the expected matching results.



References

1. Sirén K, Fischer U and Vestner J. "Automated supervised learning pipeline for non-targeted GC-MS data analysis ". Analytica Chimica Acta: X, Volume 1, March 2019.
2. S. Reichenbach, X. Tian, C. Cordero, Q. Tao. "Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography". Journal of Chromatography A, 1226:140-148, 2012.
3. S. Reichenbach, X. Tian, Q. Tao, E. Ledford, Z. Wu, O. Fiehn. "Informatics for Cross-Sample Analysis with Comprehensive Two-Dimensional Gas Chromatography and High-Resolution Mass Spectrometry (GCxGC-HRMS)". Talanta, 83(4):1279-1288, 2011.
4. Q. Tao, S. E. Reichenbach, C. Heble, and Z. Wu. "New Investigator Tools for Finding Unique and Common Components in Multiple Samples with Comprehensive Two-Dimensional Chromatography." Chromatography Today, 11, 13-18, 2018.
5. Z. Wu, J. Coleman, Q. Tao. "Distinguishing Commercial Diesel Fuel Brands Using Comprehensive Two-Dimensional Gas Chromatography/Mass Spectrometry". The International Symposium on Capillary Chromatography (ISCC), Riva del Garda, Italy, May 2018.
6. N. Hoffmann, M. Wilhelm, A. Doebbe, K. Niehaus, J. Stoye, "BiPACE 2D—graph-based multiple alignment for comprehensive 2D gas chromatography-mass spectrometry." Bioinformatics 30.7 (2014): 988-995.

Data processes and screenshots for this publication are from an alpha version of GC Image v2.9 GCxGC-HRMS (Visit www.gcimage.com for current v2.9 releases).