Advances in Instrumentation and Chemometrics of Two-Dimensional Gas Chromatography Systems

Lina Mikaliunaite

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Robert E. Synovec, Chair

Ashleigh B. Theberge

Dan Fu

Program Authorized to Offer Degree:

Chemistry

University of Washington

Abstract

Advances in instrumentation and chemometrics of two-dimensional gas chromatography systems

Lina Mikaliunaite

Chair of the Supervisory Committee:

Robert E. Synovec

Chemistry

Comprehensive two-dimensional (2D) gas chromatography (GC×GC) has gained significant popularity as an analysis tool for complex samples in metabolomics, petroleum science, and other fields. GC×GC has been shown to perform better than one-dimensional GC, including a 10-fold increase in peak capacity and more chemical selectivity. The work presented herein focuses on two aspects of GC×GC systems: advancements in instrumentation and chemometrics.

While GC×GC is used to analyze volatile compounds, when it comes to very light analytes, like permanent gasses, new instrumental designs are needed, as widely applied wall coated open tubular (WCOT) columns do not work well. Also, thermal modulators, which are in most commercially available instruments, cannot trap such low boiling point analytes due to temperature limitations. Porous layer open tubular (PLOT) columns and flow modulators must be used for such samples. However, PLOT columns are not widely used in GC×GC due to their

extreme retention on analytes, and hence being hard to pair with WCOT columns due to temperature differences that are needed for analytes to elute. In this work, we present a new GC×GC system with a PLOT column in the first dimension and a WCOT column in the second dimension. This system also used a high temperature diaphragm valve as a modulator. We showed the ability of this system to separate heavy mixtures like gasoline, and by trying thinner film PLOT columns, we showed the potential of this system to be used with even higher boiling point analytes.

GC×GC is an essential analytical technique for biological samples, as they are complex and make a perfect sample for this technique. However, due to the biological variation that is possible in these samples, data analysis can be tricky. Here, we present three projects that analyzed biological samples: cycling yeast, moisture-damaged beans, and VOCs of *Malassezia pachedermatis*. While all were very different, they were analyzed using the tile-based Fisher ratio and then post-processed to answer specific biological questions of the dataset. When analyzing cycling yeast, we were looking for analytes that were not only changing between different classes but also had a specific cycling pattern. This was achieved by simulating random data to analyze which analytes were like random patterns and which had some underlying biological information. The second dataset presented is moisture-damaged cacao beans, where the F-ratio method had to be altered, so not only the analytes that were changing with the molding process would be found, but also the hits that were bean origin-specific and did not change with molding. Lastly, we analyzed VOCs produced by *M. pachedermatis*, when grown at three different pH. In this case, the analytes need to be classified into different types, based on how much and how they changed in signal between blank media and *M. pachedermatis*, and to

be analyzed for potential pH-dependency. This was performed by post-processing using *R-metric* and *RSD-metric*.

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES xiv

ACKNOWLEDGEMENTS xvi

DEDICATION xviii

Chapter 1 . Introduction to comprehensive two-dimensional gas chromatography instrumentation and chemometric analysis 1

1.1 Overview 1

1.2 Overview of GC×GC instrument 2

1.2.1 Columns 2

1.2.1.1. Porous layer open tubular (PLOT) columns 3

1.2.1.2. PLOT columns in GC×GC systems 8

1.2.2 Modulator 8

1.2.2.1 Thermal modulators 9

1.2.2.2 Flow modulators 10

1.2.3 Detector 11

1.3 Data visualization 12

1.4 Overview of chemometric methods 13

1.4.1 Non-targeted Analysis Methods 14

1.4.1.1. Unsupervised Analysis Methods 14

1.4.1.2. Supervised Analysis Methods 15

1.4.2. Post-processing of F-ratio Results 17

1.5. Overview of the following chapters 18

1.5.1 Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection 18

1.5.2 Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension 19

1.5.3 Performance of thin-film porous layer open tubular columns in one-dimensional and comprehensive two-dimensional gas chromatography-mass spectrometry systems 20

1.5.4 Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry 21
1.5.5 Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data 22
1.5.6 Tile-based analysis with support vector machine regression modeling of GC×GC-TOFMS data of VOCs produced by Malassezia pachydermatis grown at variable pHs 23
1.6 References 24

Chapter 2 . Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection. 31
2.1 Introduction 31
2.2 Experimental 34
2.3 Results and Discussion 38
2.4 Conclusion 48
2.5 References 49

Chapter 3 . Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension .. 54
3.1 Introduction 54
3.2 Theory 58
3.3 Experimental 61
3.4 Results and Discussion 64
3.4.1 Initial Instrument Evaluation 64
3.4.2. Visualization Issues with GC×GC-qMS and Correction Method 66
3.4.3. Final Instrument Evaluation 72

3.5 Conclusion 76

3.6 References 76

Chapter 4 . Performance of thin-film porous layer open tubular columns in one-dimensional and comprehensive two-dimensional gas chromatography-mass spectrometry systems 80

4.1 Introduction 80

4.2 Experimental 82

4.2.1 Instrument parameters 82
4.2.2 Samples 85

4.3 Results and Discussion 86

4.4. Conclusion 92

4.5 References 93

Chapter 5 . Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry .. 97

5.1 Introduction 97

5.2 Experimental 101

5.3 Results and Discussion 104

5.4 Conclusion 116

5.5 References 118

Chapter 6 . Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data 122

6.1 Introduction 122

6.2 Experimental 125

6.2.1 Cacao bean samples and sample preparation 125
6.2.2 Chromatographic conditions 126
6.2.3 Data Analysis 128

6.3 Results and Discussion 132

6.4 Conclusion 142

6.5 References 143

Chapter 7 . Tile-based analysis with support vector machine regression modeling of GC×GC-TOFMS data of VOCs produced by *Malassezia pachydermatis* grown at variable pHs 148

7.1 Introduction 148

7.2 Experimental 151

7.2.1 Malassezia pachydermatis microbial culture 151
7.2.2 Device Inoculation and Incubation 151
7.2.3 HS-SPME-GC×GC-TOFMS instrument conditions 152
7.2.4 Data analysis. 153

7.3 Results and Discussion 154

7.4 Conclusion 162

7.5 References 163

Chapter 8 . Bibliography 167

Appendix A 187

Appendix B 189

Appendix C 196

LIST OF FIGURES

Figure 1.1. GG×GC instrument schematic showing the main parts of the system. 2

Figure 1.2 Schematic of wall coated open tubular columns (WCOT) on the left versus porous layer open tubular columns (PLOT). 5

Figure 1.3. Comparison between separation of 200 ppm EPA Appendix IX Volatiles Calibration mix (Sigma-Aldrich, St.Louis, MO, USA) on (A) a wall coated open tubular column (Rtx-200) and (B) a porous layer open tubular column (PLOT-Q) ran at the same chromatographic conditions, while only changing the column. 2 mL of sample were injected both times using 1:25 split, columns were run at 1.5 ml/min carrier gas, He, flow rate and analyzed using a quadruple mass spectrometer. Temperature program was stared at 100 °C and ramped at 10 °C/min to 250 °C. Lightest compounds in the mix, that separated out using PLOT-Q column, but did not using Rtx-200 column were identified in the insert of part (B). Heaviest compounds separating on both columns were identified: o-dichlorobenzene (b.p. 180 °C), m-dichlorobenzene (b.p. 173 °C), p-dichlorobenzene (b.p. 174 °C), hexachloroethane (b.p. 187 °C), 3-chloro-1,2-dibromopropane (b.p. 196 °C) and naphthalene (b.p. 218 °C). 6

Figure 1.4. Optimized separations of gasoline that was sourced from a local gas station using two different PLOT columns. (A) Separation was performed using PLOT-S column, with 0.1 mL of sample being injected with 20:1 split, where the column was operated at 1.4 ml/min flow rate of He carrier gas and analyzed by quadruple mass spectrometer. Temperature program was started at 100 °C and ramped at 15 °C/min to 250 °C. (B) Separation was performed using Rt-Alumina BOND/MAPD column, with 0.5 mL of sample being injected with 100:1 split, where the column was operated at 1.5 ml/min flow rate of He carrier gas and analyzed by quadruple mass spectrometer. Temperature program was started at 100 °C and ramped at 20 °C/min to 250 °C. Inserts in (A) and (B) show the heaviest compounds that are separated on these columns: naphthalene (b.p. 218 °C) and 1-methyl naphthalene (b.p. 244 °C). (C) and (D) show the beginning of gasoline separation on PLOT-S and Rt-Alumina BOND/MAPD columns, respectively. Isobutane and butane separate well at the beginning of the separation on both columns, but C5 components separate into 4 peaks on Rt-Alumina BOND/MAPD column, while they are overlapped on PLOT-S column (C). 7

Figure 1.5. Schematic of GC×GC modulation and data visualization in two-dimensional space. 13

Figure 1.6. Explanation of the tile-based Fisher ratio (F-ratio), four tiling schemes, and "pinning and clustering". 16

Figure 2.1. (A) Schematic of the major components of DPGM-GC×GC-FID instrument. A pulse valve provides dynamic pressure gradient modulation (DPGM) between the ^{1}D and ^{2}D columns, at the T-union by applying a suitable pressure (P_{aux}). Detection was performed with a flame ionization detector (FID). (B) Background chromatogram collected with the temperature program from 200ºC to 250 ºC at 10 ºC /min. 35

Figure 2.2. Background noise for four consecutive runs of solvent (acetone) chromatogram (blue) and instrument background (air injection) chromatogram (red). The subtraction of the

instrument background from the solvent chromatogram is shown in black for runs 1-4. In blue, the acetone blank is shown, and in red, the instrument background is shown. In black, the subtraction of the instrument background from the acetone blank is presented. 38

Figure 2.3. Demonstration of the baseline correction method. (A) Raw (preprocessed) modulated chromatogram of the 500-ppm alkane standard solution: (1) pentane, (2) hexane, and (3) heptane. (B) Raw background chromatogram (no sample injected). (C) Chromatogram after the background subtraction step. (D) Chromatogram after the Savitzky-Golay smoothing step. The insert shows the zoomed-in view of the noise between 6 to 7 min. (E) Standard deviation of the baseline noise in the 6 to 7 min window following background subtraction as a function of the normalization factor (N_f) applied to the background in (B) prior to subtraction. The N_f corresponding to the lowest standard deviation, is the optimized N_f. (F) Fully processed chromatogram when optimized N_f is used ($N_f = 1.02$). 40

Figure 2.4. (A) Raw modulated chromatogram of the 31-ppm alkane standard solution: (1) pentane, (2) hexane, and (3) heptane. (B) Raw background chromatogram (no analyte injected). (C) Chromatogram after the background subtraction step. (D) Chromatogram after the Savitzky-Golay smoothing step when an optimized normalization factor ($N_f = 1.02$) is used. The insert shows the zoomed-in version of the two modulated ^{2}D hexane peaks. 42

Figure 2.5. Overlay of processed vector chromatograms of 31-ppm alkane standard solution produced using different normalization factors (±10% from $N_f = 1$), where the black line represents using $N_f = 1$, and the green line represents the optimal $N_f = 1.02$. (A) Baseline noise section (no analytes eluting). (B) Section of the data where hexane ^{2}D peaks elute. (C) Unfolded data of 31-ppm hexane standard solution for 4 replicates, after data processing using an optimized N_f ranging from 1.00 to 1.02. (D) Summed ^{2}D hexane peaks obtained by summing along ^{1}D for 4 replicates at 31 ppm and 7.8 ppm. 44

Figure 2.6. (A) Peak area versus concentration calibration plot for hexane (7.8 ppm – 4000 ppm). Peak areas measured from summed ^{2}D peaks as in Fig. 5D. Calibration plots for lowest three concentrations (7.8 ppm, 15.6 ppm, 31.3 ppm) with optimal N_f (B) and with $N_f = 1$ (C). (D) Percent deviation (% deviation) between peak areas obtained when N_f = optimal and when N_f = 1 for the full concentration range 7.8 ppm – 4000 ppm on a $\log_2$ concentration scale. 45

Figure 2.7. Analysis of a natural gas sample. (A) Unfolded vector form chromatogram of a natural gas sample separation, following the baseline correction method, showing full separation of light hydrocarbons in the sample. (B) Processed GC×GC chromatogram that has made from (A) by folding the data at the PM. Methane peak was purposely plotted on a 10-fold less sensitive scale, to more accurately reveal less concentrated peaks in the sample. (C) The modulated isopentane 2D peaks with the signal enhancement (SE) calculation illustrated. The insert shows the zoom-in of the tallest 2D peak. (D) Unmodulated isopentane 1D peak (red) with a white trace showing the unmodulated peak that has been boxcar averaged to contain the same number of data points as the tallest modulated 2D peak in (C). In both (C) and (D) the retention times shown are relative; the pentane peak in (C) elutes about 2 min later than in (D). 47

Figure 2.8. Analysis of an unleaded gasoline sample. GC×GC chromatograms are produced for the separation at (A) $P_M = 2$ s, $p_w = 0.5$ s and (B) $P_M = 6$ s, $p_w = 2$ s, with the baseline correction

method applied in both cases. (C) GC×GC chromatogram is produced for the separation at P_M = 6 s, p_w = 2 s without applying the baseline correction method. After 25 min, darker feature is due to the baseline noise. 48

Figure 3.1. (A) Schematic of the major components of the GC×GC-qMS instrument. A high temperature diaphragm valve (HT-DV) with a 10 µL sample loop is used as a modulator between the first-dimension (1D) and second-dimension (2D) columns. Detection was performed with a quadrupole mass spectrometer (qMS). 55

Figure 3.2. (A) Plot of scanned voltage versus time for the qMS in the context of Eqs. (1-3). The cycle time is defined as the time between starting two consecutive scans. The scan time is amount of the time that the voltages (DC and RF) are linearly ramped, and the interscan delay is the time the voltage is held constant between scans. (B) Relationship between the mass range and cycle time is shown for every mass range from 2 to 350 mass channels (m/z 45 to m/z 394) for the specific qMS instrument that was used. Insert shows relationship between scanned mass channel range and the scan rate, which is calculated by taking an inverse of cycle time. 59

Figure 3.3. (A) Unfolded data for a simulated analyte at a scan rate of 20.00 Hz and a P_M of 2 s, which produces 40 data points (pts) per P_M. (B) Reshaped data for the analyte in (A). Here, the cycle time is divided evenly into the P_M. (C) Unfolded data for an analyte simulated at a scan rate of 21.28 Hz and P_M of 2 s, showing the variation of number of data points for every modulation (42-43 data points per P_M). (D) Reshaped chromatogram for the analyte in (C). Here, the cycle time does not divide evenly into the P_M, thus distorting the peak. 60

Figure 3.4. (A) Total ion current (TIC) chromatogram of the HT-DV modulated GC×GC-qMS separation of gasoline, where several alkanes (C4-C9) and some aromatics are identified. Isobutane and butane peaks were purposely plotted on a 10-fold more sensitive signal scale, to reveal these less concentrated peaks. (B) Zoom-in between 8 and 11 min of (A), showing various aromatic compounds readily separated using the PLOT-S × Rtx-200 column configuration. (C) The two first peaks separated in gasoline, isobutane and butane, are shown in an unfolded raw-data chromatogram, where circles on the trace show the positions of the mass spectrum scans (data points). 65

Figure 3.5. TIC chromatograms of the HT-DV modulated GC×GC-qMS separation of gasoline when different scan rates are used, and data is simply folded into the 2D space using an integer number of points in the second dimension (40 points for 2 s P_M), showing the shift of peaks. (A) GC×GC chromatogram collected at 20.01 Hz. (B) Inset of unfolded butane peak at 20.01 Hz. The first modulation is shown in blue, the second in red, and the third in pink. (C) Illustration of the amount of $t_{shift,mod}$ at 20.01 Hz. Each colored dot shows the ^2t_r for each ^{2}D peak and the measured $t_{shift,mod}$ is illustrated. (D) GC×GC chromatogram collected at 19.88 Hz. (E) Inset of unfolded butane peak at 19.88 Hz. (F) Illustration of the amount of $t_{shift,mod}$ at 19.88 Hz. (G) GC×GC chromatogram collected at 20.20 Hz. (H) Inset of unfolded butane peak at 20.20 Hz. (I) Illustration of the amount of $t_{shift,mod}$ at 20.20 Hz. 67

Figure 3.6. Plots showing errors associated with different scan rates using the qMS instrument. (A) The $t_{shift,mod}$ is calculated using Eq. (4) for every scan rate, produced when using mass ranges from 2 to 350 mass channels (m/z 45 to m/z 394). (B) The $t_{shift,total}$ is calculated in terms of how

many modulations of shift using Eq. (5),using $P_M = 2$ s. The red line indicates the ½P_M of shift per 15 min run. (C) At this threshold the zoom-in of the graph is provided. A total of 44 scan rates fall under this threshold, out of which 6 have a shift of 0. These 44 scan rates are listed in Table 2. 68

Figure 3.7. Computational resampling method solution to using a cycle time that does not divide evenly into the P_M. (A) Raw data collected at 20.20 Hz. (B) Literature method of resampling the vector of data into the closest integer scan rate. (C) Reregistered ^{2}D chromatogram to correct wraparound of aromatic compounds. (D) Zoom-in of a section between 7.5 and 8.5 min for the resampled and reregistered data, highlights that resampling provides non-distorted 2D peak shapes. 71

Figure 3.8. Separations of Calibration MegaMix using HT-DV modulated GC×GC-qMS with (A) 13 m and (B) 5 m (B) PLOT-S columns. Experimental details for both runs are provided in Table 1. For both separations, the beginning of each chromatogram was plotted on a more sensitive scale to better show less concentrated compounds. All analytes that were identified with numbers are listed in Table 3.3. For analytes 8, 12 and 53, that had very small signal compared to the other peaks, a dot indicates their peak location. 72

Figure 3.9. ^{2}D summed peak profiles used to evaluate peak height and area reproducibility for three analytes in the JP8 fuel using m/z indicated showing overlay of 9 injection replicates. (A) Trans-decalin (m/z 138) with a peak height RSD of 6.7 % and area RSD of 7.7 %. (B) 2-dimethylsterene (m/z 132) with a peak height RSD of 13.2 % and area RSD of 12.5 %. (C) Azulene (m/z 128) with a peak height RSD of 7.7 % and area RSD of 7.7 %. 75

Figure 3.10. 2D summed peak profiles used to evaluate the peak height based LOD for three non-native analytes spiked at 50 mg/L in the JP8 fuel using m/z indicated showing overlay of 3 injection replicates. (A) 1-chlorohexane (m/z 93) has a LOD of 3.4 mg/L. (B) α-pinene (m/z 136) has a LOD of 7.0 mg/L. (C) Bromobenzene (m/z 158) has a LOD of 2.1 mg/L. 75

Figure 4.1. One-dimensional separations of Calibration MegaMix using (A) 8 µm film thickness column, (B) 4 µm film thickness column, and (C) 2 µm film thickness column. Chromatograms of different hues but the same color are replicate separations performed using the same thickness column that was manufactured at different times. (D) Temperature program for the separations shown in A-C. 86

Figure 4.2. One-dimensional separations of the lightest 5 analytes in Refinery Mix using (A) 8 µm film thickness column, (B) 4 µm film thickness column, and (C) 2 µm film thickness column. The five analytes numbered are: 1 – nitrogen, 2 – methane, 3 – carbon dioxide, 4 – ethene, and 5 – ethane. Chromatograms of different hues but the same color are replicate separations performed using the same thickness column that was manufactured at different times. (D) Temperature program for the separations shown in A-C. 87

Figure 4.3. One-dimensional separations of Refinery Mix using (A) 2 µm film thickness PLOT-S column and (B) Rxi-5 wall coated column. 89

Figure 4.4. Two-dimensional separations of Calibration Megamix using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first

dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B. 90

Figure 4.5. Two-dimensional separations of Refinery Mix using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B. 91

Figure 4.6. Two-dimensional separations of gasoline using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B. 92

Figure 5.1. (A) GC×GC-TOFMS chromatogram of yeast metabolite sample using m/z 73, where the chromatogram on ^{1}D starts at 5 min, to account for the solvent delay. (B) A zoomed-in portion of the chromatogram to facilitate visualizing the separation. 104

Figure 5.2. Examples of the metabolite cycling patterns in the context of the changes in dissolved O2 level over the 2 cycles (5 hr per cycle period) for the diploid yeast strain CEN.PK. (A) Myo-inositol exhibits a 1 cycle/5 hr cycle sinusoidal pattern in its 24 time-point signal sequence. (B) Methyl citrate exhibits a “spiky” cycling pattern, where there are two high intensity signals separated by a 12 time-point interval (exactly 1 cycling period). (C) Glucose-6-phosphate exhibits multimodal cycling, where it will be shown there is a correlation between the first half (time-points 1 – 12) and the second half (time-points 13-24) of the 24 time-point signal sequence. (D) Relative O2 concentration oscillation pattern, where the 24 black dots indicate when the yeast samples were collected (25 min apart). 105

Figure 5.3. The process for calculating the RSD and LOF metrics to find cycling metabolites is illustrated by examples. (A) Myo-inositol 24 time-point signal sequence with an RSD of 40.3%. (B) Determination of LOF for myo-inositol where the first 12 time-points (1 – 12) and the last 12 time-points (13 – 24) from (A) are matched up 1 to 13, 2 to 14, etc., and then Eq. (1) is applied, resulting in a LOF of 22.0%. (C) Hexanoic acid 24 time-point signal sequence with an RSD of 40.9%. (D) Determination of the LOF for hexanoic acid resulting in a LOF of 51.5%. (E) Randomly generated signal sequence with an RSD of 40.3%. (F) Determination of the LOF for the randomly generated signal sequence in (E) resulting in a LOF of 70.2%. 106

Figure 5.4. Steps in the computational method to identify cycling metabolites. (A) Overlay of the observed 2D distributions of LOF versus RSD, for three simulated RSD levels: 15%, 35% and 75%. There are 10000 simulations at each simulated RSD level. (B) Frequency of the LOF obtained for the RSD simulated at 15%, 35%, 55%, 75% and 95%. (C) Overlay of all 2D distributions of LOF versus RSD for the randomly generated signal sequences. There were 10000 randomly generated signal sequences created at each RSD from 10% to 100% spaced every 5%, resulting in 19 simulated RSD conditions or a total of 190,000 sequences. The “slicing method” is illustrated at an RSD of 75% ±2.5%, which contained 8996 LOF versus RSD values. The LOF threshold at the 95% confidence level ($p < 0.05$) is indicated by the transition from blue to red dots. (D) Frequency of LOF for each centered RSD level as illustrated for RSD of 75% in (C). (E) Illustration of the 95% confidence level for the RSD centered at 75% in (C). (F)

Application of the 95% confidence level generated in (C) to the metabolite signal sequences. The 672 metabolite hits discovered by F-ratio analysis divide into two groups: (red dots) 210 metabolites that exhibit non-random (cycling) behavior, and (blue dots) 462 metabolites that are not cycling. 107

Figure 5.5. Metabolite cycling phase determination process. (A) The 1 cycle/5 hr period sine wave model (red triangles) of 24 time-points (2 cycles over 10 hr) was constructed with the same SD as the signal sequence for myo-inositol (blue dots). (B) The LOF between the sine wave model and myo-inositol signal sequence was repeatedly calculated via Eq. (1) by moving the sine wave one time-point over, i.e., 30° phase change per time-point, with an RSD_{LOF} of 38.29% obtained. (C) A randomly generated signal sequence (blue dots) is shown in relation to a sine wave model (red triangles) with the same SD. (D) The LOF between the sine wave model and randomly generated signal sequence is shown, with an RSD_{LOF} of 11.64% obtained. 110

Figure 5.6. Statistically based cycling and phase assignment for the 1 cycle/5 hr period and 2 cycle/5 hr period metabolites. (A) Distribution of RSD_{LOF} for the 190000 randomly generated signal sequences using the 1 cycle/5 hr period sine wave model (eg., Fig. 5.5. D). The 95% confidence level ($p < 0.05$) at an RSD_{LOF} of 19.16% was determined (red line). (B) Distribution of RSD_{LOF} for the 210 metabolite signal sequences (eg., Fig. 5.5. B) in which 104 exceed the RSD_{LOF} of 19.16% threshold. (C) Distribution of RSD_{LOF} for the 190000 randomly generated signal sequences using a 2 cycle/5 hr period sine wave model. The 95% confidence level ($p < 0.05$) at an RSD_{LOF} of 20.19% was determined (red line). (D) Distribution of RSD_{LOF} for the 210 metabolite signal sequences in which 13 exceed the RSD_{LOF} of 20.19% threshold.....112

Figure 5.7. Heatmap for the 55 identified metabolites (per Table 5.1.), arranged by cycling category: first 35 metabolites are 1 cycle/5 hr period (1c), separated by white line from D-ribulose 5-phosphate which is the only identified metabolite of the 2 cycles/5 hr period (2c) type, separated with a white line from 5 spiky (sp) metabolites that are separated by another white line from the remaining 14 multimodal (mm) metabolites. Every category except multimodal is ordered according to phase. The red color indicates a high concentration and blue identifies a low concentration. Results for the 155 unknown metabolites are provided in Table B.2. and Fig. B.4. 115

Figure 5.8. (A) PCA scores plot for all 210 cycling metabolites labeled by cycling category: 1 cycle/5 hr period (purple), 2 cycles/5 hr period (yellow), spiky (red), or multimodal (green). (B) PC2 loadings, showing a repeating cycling pattern between first 12 time-points and last 12 time-points. 116

Figure 6.1. GC×GC total ion current (TIC) chromatograms of beans from two of the six geographical regions and from different molding days. Beans from (A) Panama and (B) Venezuela on Day 1 of the molding process. Then, beans from (C) Panama and (D) Venezuela on the Day 6 of molding. Chromatographic 2D locations of Hit 1 (3-methyl-2-nonene) and Hit 15 (2-methyl-3-buten-2-ol) are labeled on each chromatogram. 126

Figure 6.2. Selective ion chromatograms using the signal from m/z 70, which yielded the highest F-ratio for Hit 1 (3-methyl-2-nonene), a kinetic hit. For Day 1, 3-methyl-2-nonene has a high concentration both in (A) Panama and (B) Venezuela beans, however throughout the molding

process, specifically by Day 6, 3-methyl-2-nonene almost completely disappears in both (C) Panama and (D) Venezuela beans. 127

Figure 6.3. Selective ion chromatograms made using m/z 50, which yielded the highest F-ratio for Hit 15 (2-methyl-3-buten-2-ol), which is a geographical region hit. On Day 1, 2-methyl-3-buten-2-ol is seen to be present in Panama beans for (A) Day 1 and (B) Day 6, however, 2-methyl-3-buten-2-ol is not visually present in Venezuela beans for either (C) Day 1 (C) or (D) Day 6. 128

Figure 6.4. Composite chemical fingerprints (CCFs) for (A) 3-methyl-2-nonene (Hit 1) and (B) 2-methyl-3-buten-2-ol (Hit 15). The 36 signals in each CCF have been scaled to the highest signal, where dark red represents the highest relative signal of 1, and dark blue represents the lowest relative signal of 0. Bean type labels: Panama (P), Costa Rica (CR), Ivory Coast (IC), Ghana (G), Ecuador (E), and Venezuela (V). (A) 3-methyl-2-nonene has high signal only for Day 1, producing an F-ratio$_{kinetic}$ of 279.4 and F-ratio$_{region}$ of 0.01. (B) 2-methyl-3-buten-2-ol has relatively high signals for all of the molding days in the beans from the Panama region as well as Ivory coast, resulting in an F-ratio$_{region}$ of 63.5 and an F-ratio$_{kinetic}$ of 0.2. (C) An example of when the signals in these CCFs are randomized for 3-methyl-2-nonene, producing a null CCF pattern resulting in an F-ratio$_{kinetic}$ of 1.3 and F-ratio$_{region}$ of 0.5. (D) Similarly, an example of randomizing the signals for 2-methyl-3-buen-2-ol, producing a null CCF pattern resulting in an F-ratio$_{kinetic}$ of 1.4 and F-ratio$_{region}$ of 0.9. 130

Figure 6.5. F-ratio distributions for (A) F-ratio$_{kinetic}$ and (B) F-ratio$_{region}$ values using log scale. (C) Distribution for F-ratio values in log scale for the 5.9×10^{6} randomized null CCFs. (D) Various percentile thresholds for the pseudo-null distribution at 99%, 99.9%, 99.99% and 99.999% using a zoomed-in distribution for F-ratio values in log scale for the randomized null CCFs. 136

Figure 6.6. (A) Relationship between F-ratio$_{kinetic}$ and F-ratio$_{region}$ for every hit presented in log scale. Hits with a higher F-ratio$_{kinetic}$ are in blue and the hits with a higher F-ratio$_{region}$ are in light red. (B) Relationship between the highest F-ratio values from (A) in log scale and hit number. Various thresholds are marked with a dashed line, with the darkest color denoting the strictest threshold of 99.999%, and the lightest color the threshold of 99%. 137

Figure 6.7. PCA scores plots for (A) all the hits exceeding the 99% null threshold (Th) using 2 principal components (PCs), (B) when using 3 PCs. PCA scores plots for (C) only the kinetic hits exceeding the 99% null threshold and (D) only the region hits exceeding the 99% null threshold. Kinetic hits are in blue shaded circles and region hits are in red shaded circles. The darkness of color shade represents what null distribution threshold the hit was found 99.999% being represented by the darkest color shade and 99% by the lightest color shade, with 99.99% and 99.9% being one hue lighter and two hues lighter, respectively. 138

Figure 6.8. PCA scores plot of the kinetic hits that passed the strictest 99.999% null threshold. Also indicated are the color-coded composite chemical fingerprints (CCFs) for seven example analyte hits (as in Fig. 6.4). 139

Figure 6.9. PCA scores plot of the region hits that passed the strictest 99.999% null threshold. Also indicated are the CCFs for seven example analytes. 140

Figure 7.1. Zoom-in of selective ion chromatograms for (A) blank at pH 5, (B) blank at pH 9, (C) blank at pH 12, (D) Malassezia at pH 5, (E) Malassezia at pH 9, (F) Malassezia at pH 12 of (i) hit 1 at m/z 116, (ii) hit 2 at m/z 107 and (iii) hit 29 at m/z 60. (G) summed signal profile for (i) hit 1, (ii) hit 2 and (iii) hit 29. Error bars in G i-iii represent standard error of the mean of five replicates. 153

Figure 7.2. Original signal profiles for (A) hit 1, (B) hit 2, and (C) hit 29. Explanation of calculations of (D-F) R value, (G-I) RSD, and (J-L) LOF. Error bars in A-C represent standard error of the mean of five replicates. 155

Figure 7.3. (A) Relationship between R, RSD, and the log of LOF for 566 analytes found through tile-based F-ratio analysis. (B) Relationship of R and LOF for 566 analytes. (C) Relationship of R and RSD for 566 analytes. (D) PCA plot made using signal profiles of 566 analytes. Color coding represents the class that the analyte was assigned based on its R value, where pink are analytes being consumed, blue are pH dependent analytes and red are analytes that are being produced by Malassezia. 158

Figure 7.4. Plot of R vs RSD for 566 analytes, showing signal profile examples for analytes at similar R values, but different RSD values for three different classes of analytes. Error bars in signal patterns represent standard error of the mean of five replicates..... 160

Figure 7.5. Signal profiles of top 20 identified analytes of each class: (A) for consumed analytes, (B) for analytes that are pH dependent and (C) for produced analytes. 161

Figure A.1. Sample chromatogram of JP8 fuel indicating the location of three native analytes (black dots) used for peak reproducibility study (Fig. 3.9.) and location of three spiked non-native compounds (red dots) used for LOD study (Fig. 3.10.). 188

Figure B.1. Averaging 72 peak areas to 24 time-point signal sequences. Using the top F-ratio m/z for each of the 672 hits, the three replicates at each time-point were averaged to reduce the data density from 72 points in (A) to the final 24-point signal sequences in (B) for each of the hits (myo-inositol shown as an example). The averaged 24-point signal sequence is then used in all subsequent data analysis steps. 190

Figure B.2. To provide a robust and accurate determination of the LOF cycling pattern threshold, randomly generated signal sequences span a range of RSD to match the range of the experimental GC×GC-TOFMS data (both normalized to have a mean of 1). Here myo-inositol is used as an example to show that normalization of the data to the mean has no impact over LOF, where in (A) and (B) the LOF is calculated for non-normalized data, for comparison to (C) and (D), where the normalized peak areas are used. In (B) and (D), the LOF calculation is illustrated for the LOF determination for myo-inositol where the first 12 time-points (1 – 12) and the last 12 time-points (13 – 24) from (A) or (C) accordingly are matched up, 1 to 13, 2 to 14, etc., and then Eq. (1) is applied to result in the same LOF of 22.0%..... 191

Figure B.3. LOF versus RSD statistical threshold determination was applied from a simulated RSD range of 10% to 100%, the resulting 95% confidence interval for LOF were fit to an equation and extrapolated to slightly higher RSD using Matlab 2019b using Curve Fitting Toolbox. The polynomial model, as well as goodness-of-fit parameters are also provided. 192

Figure B.4. Heatmap of the 155 unknown metabolites ordered according to Table B.2., where the signal for each time interval is shown where white lines are separating different types of metabolites; lowest concentration is in blue and highest concentration is in red. 194

Figure B.5. PC 1 (A) and PC3 (B) loadings, showing a somewhat repeating cycling pattern between the first 12 time-points and the last 12 time-points. 195

Figure C.1. TIC chromatograms of blank at (A) pH 5.7, (B) pH 9.7, (C) pH 12.4, and of M. pachydermatis at (D) pH 5.7, (E) pH 9.7, (F) pH 12.4. 196

Figure C.2. Zoom-in of TIC chromatograms of blank at (A) pH 5.7, (B) pH 9.7, (C) pH 12.4, and of M. pachydermatis at (D) pH 5.7, (E) pH 9.7, (F) pH 12.4. 197

LIST OF TABLES

Table 1.1. Commercially available porous layer open tubular (PLOT) columns of different stationary phases (porous layers) compared between Restek, Supelco, Agilent and Quadrex. Table remade and updated from Restek's PLOT Family document from 2020. 3

Table 3.1. Experimental parameters for the studies performed. 62

Table 3.2. Table of scan rates that produce less than half of the PM of shift, along with the scan rate, the mass range that corresponds to the scan rate and the number of modulations of shift in a 15 min GC×GC-qMS run calculated using Eq. (5) and converted to units of number of modulations. 69

Table 3.3. Identified compounds and their boiling points in the Calibration MegaMix, with the peaks labeled by identification (ID) number in Fig. 3.8. Retention time entries of "N/A" indicate that analyte was not observed. 72

Table 4.1. Experimental parameters for the one-dimensional comparison studies performed between different film thickness PLOT columns. 83

Table 4.2. Experimental parameters for the one-dimensional comparison study performed between the PLOT column and WCOT column. 84

Table 4.3. Experimental parameters for the two-dimensional comparison study performed between the different film thickness PLOT columns in the first dimension. 85

Table 5.1. The 55 identified metabolites were arranged according to their cycling category: 1 cycle/5 hr period (1c), 2 cycles/5 hr period (2c), spiky (sp), or multimodal (mm). A star (*) near the analytes indicates that were previously found [30]. For every cycling category except multimodal, the metabolites were ordered according to phase in degrees. An accompanying heatmap is provided in Fig. 5.7. 113

Table 7.1. Summary of results of support vector machine regression model using cross-validation results. 162

Table A.1. Summary of the experimental parameters used in peak reproducibility and LOD study, where JP8 fuel was analyzed using WCOT columns on both dimensions. 187

Table B.1. Summary of the parameters used in the implementation of ChromaTOF Tile Software version 101 (LECO, St. Joseph, MI, USA) to perform tile-based F-ratio analysis, and quantify each metabolite hit in the hit list at each time-point. 189

Table B.2. Final hitlist for the 155 unknown metabolites, where metabolites are ordered by category: 1 cycle/period (1c), 2 cycle/period (2c), spiky (sp) and multimodal (mm) and within each category is ordered by increasing phase (degrees). 193

Table C.1. Standard parameters used to build for support vector machine regression model. ... 198

Table C.2. The final hitlist of 566 analytes, with their identification and Fisher ratio (F-ratio). 199

ACKNOWLEDGEMENTS

There are many people without whom this dissertation would not have been possible. First, I would like to thank my Ph.D. advisor, Dr. Robert Synovec, for supporting me throughout many projects, all the ups and downs of my Ph.D., and helping me develop as a scientist. While I learned a lot in the areas of chromatography and chemometrics in his group, most importantly, I fell in love with GC×GC as an analytical technique and will forever be thankful for the opportunity to learn about it. I also want to thank my advisor for sending me to many national and international conferences, where I got to interact with many scientists in my field and make valuable connections. One of these connections I want to thank is Dr. Katelyn Perrault, who made me feel very welcome at my very first Multidimensional Chromatography Workshop and since has been a great mentor. I also want to thank Dr. František Tureček for sending my acceptance to UW letter and being part of my second-year committee. I also want to thank my committee for your time and all the guidance you gave me during my general exam and to my dissertation defense day. Further, I would like to thank my undergraduate advisor, Dr. David Green. He is responsible for making me fall in love with chromatography, and since the first LC chromatogram I ran in his lab I was sold on chromatography and never looked back. Thank you for encouraging me to apply to graduate school and making many impossible things happen so I could get my bachelor's degree from Pepperdine. Most importantly, thank you for always wanting to hear how graduate school was going and telling me that it would all be ok.

Second, I want to thank the current Synovec group members for all the support and fun. In particular, I want to thank Wenjing Ma for being a great friend and someone I could always trust and talk to when research got frustrating. I cannot thank two alums of Synovec group members enough, Dr. Paige Sudol and Dr. Caitlin Cain, for training me in the lab and answering

many questions I always had. I also want to thank Jamison Whitten and Jodie Tokihiro for all of your help with our microbial project – it took a lot of time and effort, and it was great doing it with you both.

Then, I want to thank my family, Mom Eglė, Dad Šarūnas, and Grandma Vida, for lots of encouragement, FaceTime calls, and for coming all the way to Seattle to see me once in a while. Thank you for believing my 18-year-old self when I asked you to let me study an ocean and continent away from home. Finally, I want to thank my now husband Jordan Ehrman, for being my best friend. The University of Washington has not only given me a Ph.D., but it also gave me you, and due to that, it will always hold a special place in my heart. Since I met you in early 2020 at UW visitation weekend, you have become my best friend, my biggest supporter, and someone who told me that hard work and tears would all be worth the Ph.D. I love you for that, and I cannot be happier that we got to have this graduate school experience together.

DEDICATION

To my mom, Eglė, and dad, Šarūnas, for always supporting and believing in my big dreams, and to my grandma, Vida, who sacrificed her sleep my entire childhood, reading textbooks, just to assist me in my learning.

Chapter 1. Introduction to comprehensive two-dimensional gas chromatography instrumentation and chemometric analysis

This chapter is reproduced from Mikaliunaite, L., Bell, D. S., Synovec, R. "Application of Porous Layer Open Tubular Columns: Beyond Permanent Gases" *LCGC North America*, *40* (10) and Trinklein, T. J., Cain, C. N., Ochoa, G. S., Schöneich, S., Mikaliunaite, L., & Synovec, R. E. "Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis" *Analytical Chemistry*, *95*(1), 264–286.

1.1 Overview

The work described in this dissertation covers advances in instrumentation and data analysis of comprehensive two-dimensional gas chromatography (GC×GC) systems. Two-dimensional gas chromatography (GC×GC) was pioneered by Liu and Philips in 1991 [1], and since then has grown in popularity as an analysis technique for complex volatile and semi-volatile analyte mixtures [2–4]. It is an indispensable technique in the analysis of fuels [5,6], biological samples [7,8], and others [9]. Compared to one-dimensional gas chromatography (GC), GC×GC can achieve about 10 times higher peak capacity and provide increased detectability [10]. In GC×GC systems, standard GC separation is performed on the first-dimension column (^{1}D), which is then sampled in small amounts throughout the ^{1}D separation (every 1-3 seconds) using a modulator and reinjected onto the second-dimension column (^{2}D) for a complimentary separation. When GC×GC is paired with multivariate detectors like a time-of-flight mass spectrometer (TOFMS) or a high-resolution mass spectrometer (HRMS), it allows for fast screening of samples, with easy identification, and the possibility to use complex chemometric techniques to model different classes of samples and answer complex analytical questions.

1.2 Overview of GC×GC instrument

The schematic of a standard GC×GC instrument is shown in Figure 1.1. The flow path of the instrument starts at the inlet, which is the same as the standard one-dimensional GC instrument. Then it is preceded by a first-dimension (^{1}D) column, which is connected to a second-dimension (^{2}D) column by a modulator. There are a couple of commercially available GC×GC instruments like LECO's Pegasus BT 4D or Pegasus BTX. However, many one-dimensional GC instruments can be retrofitted to be able to perform GC×GC, by building them in a modulator like a high-temperature diaphragm valve [11] or one of many commercially available options, like SepSolve's INSIGHT- modulator [12].

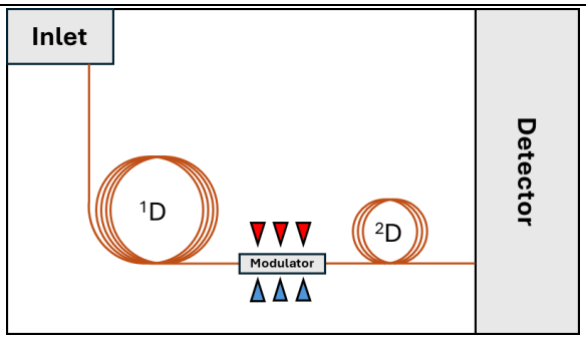


Figure 1.1. GG×GC instrument schematic showing the main parts of the system.

1.2.1 Columns

Generally, in a GC×GC instrument, the first-dimension column is a longer (~30-60 meters long) column, while the second-dimension column is a shorter (~1-5 meters long) column. The columns not only vary in length but also usually have complimentary phases to one another [13]. If the ^{1}D column is nonpolar, and the ^{2}D column is more polar than the ^{1}D column, that type of separation is called normal column configuration. This type of separation will have alkanes at the bottom of the 2D separation time (low retention on ^{2}D column), and aromatics at the top of the 2D separation time (high retention on ^{2}D column). It is generally used in samples where the interest is to investigate the alkanes, as they are more clearly separated using normal

column configuration [12]. On the other hand, if the ^{1}D column is polar, and the ^{2}D column is more nonpolar than the ^{1}D column, that type of separation is called reverse column configuration. This type of separation will have aromatics at the bottom of the 2D separation time (low retention on ^{2}D column), and alkanes at the top of the 2D separation time (high retention on ^{2}D column). It is used for fuel analysis, as the main interest for fuels is usually in the aromatic compounds [6]. Generally, most GC×GC set-ups use widely applicable wall coated open tubular (WCOT) columns, however a less explored type of column to use for GC×GC is porous layer open tubular (PLOT) columns.

1.2.1.1. Porous layer open tubular (PLOT) columns

Porous layer open tubular (PLOT) columns are traditionally built with particles that are adhered to the tubing walls, while WCOT columns have a stationary liquid phase layered on the wall of the capillary (Figure 1.2). There are a wide variety of commercially available stationary phases of porous layer open tubular (PLOT) columns (Table 1.1.) such as alumina, molecular sieve, carbon, silica, and a variety of porous polymers.

Table 1.1. Commercially available porous layer open tubular (PLOT) columns of different stationary phases (porous layers) compared between Restek, Supelco, Agilent and Quadrex. Table remade and updated from Restek's PLOT Family document from 2020.

Porous Layer	Restek	Supelco	Agilent	Quadrex
Bonded Silica	------	------	CP Silica PLOT, GS-GasPro	------
Aluminum oxide	Alumina BOND/Na_2SO_4	Alumina-Sulfate	GS-Alumina, CP-Al_2O_3/Na_2SO_4	------
Aluminum oxide	Alumina BOND/KCl	Alumina-Chloride	GS-Alumina KCl, HP PLOT Al_2O_3, CP-Al_2O_3/KCl	PLT-AL2O3
Aluminum oxide	Alumina BOND/CFC	------	------	------
Aluminum oxide	Alumina BOND/MAPD	------	Select Al_2O_3 MAPD	------

Molecular Sieve 5A	Msieve 5A	Molsieve 5A	HP PLOT Molesieve, CP-Molesieve 5A	PLT-5A
Carbon	------	Carboxen-1010 PLOT, Carboxen-1006 PLOT	CarboBOND, CarboPLOT P7, GS-CarbonPLOT	------
100% Divinylbenzene	Q-BOND	Supel-Q-PLOT	HP PLOT Q, CP-PoraPLOT Q, CP-PoraBOND Q	PLT-Q
Intermediate polarity porous polymer	QS-BOND	------	GS-Q	------
DVB vinylpyridine polymer	S-BOND	------	CP-PoraPLOT S	------
DVB ethylene glycol-dimethylacrylate polymer	U-BOND	------	HP PLOT U, CP-PoraPLOT U, CP-PoraBOND U	PLT-U

Alumina columns have an aluminum oxide stationary phase that is then deactivated with various salts. These columns have high capacity and produce symmetric peaks. Molecular sieve columns are designed specifically for permanent gas separations. Carbon-layer columns are made to have selectivity for inorganic and organic gases, like separating CO and CO_2. Silica columns are a great alternative to packed columns, as they are made to be selective for C1-C4 compounds. Porous polymer columns are hydrophobic which makes them more applicable to the analysis of wider range of samples. Porous polymer columns come in different polarities, with divinylbenzene being the most nonpolar, intermediate polarity porous polymer and DVB vinylpyridine polymer are mid-polar, and divinylbenzene ethylene glycol/dimethylacrylate is the most polar polymer column. These commercially available PLOT columns (Table 1.1.) are extensively tested to produce superior separations and are made to not shed particles, which previously was a concern with PLOT columns [14].

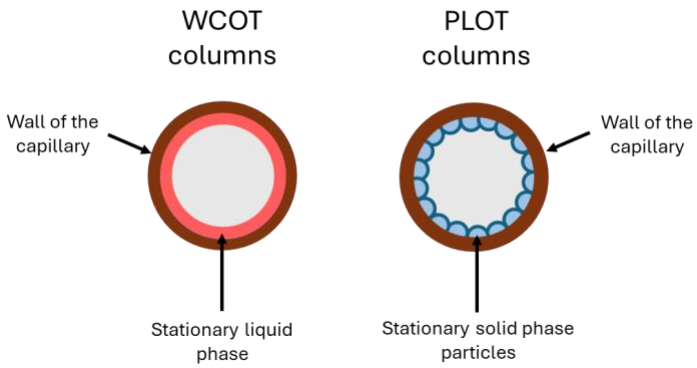


Figure 1.2 Schematic of wall coated open tubular columns (WCOT) on the left versus porous layer open tubular columns (PLOT).

The first lab-made PLOT columns were introduced in 1963 [15,16]. Since then, these columns have been mostly used for the separations of light mixtures, like permanent gases [17–19] or light hydrocarbons [20–22]. It has been shown that PLOT columns are very retentive for light compounds. For example, a 10-meter molecular sieve column has been used to separate permanent gases in the presence of hydrocarbons and nitriles [17]. With exception of He and Ne, which coeluted, all the other permanent gases (Ar, N_2, Kr, Xe) as well as methane and carbon monoxide were baseline separated [17]. Alumina columns have been evaluated by Ji et al., where authors showed that it is highly retentive for C1-C6 compounds [20]. Baseline separation of methane, ethane, ethylene, propane as well as higher carbon content compounds was shown using 30-meter alumina column [20].

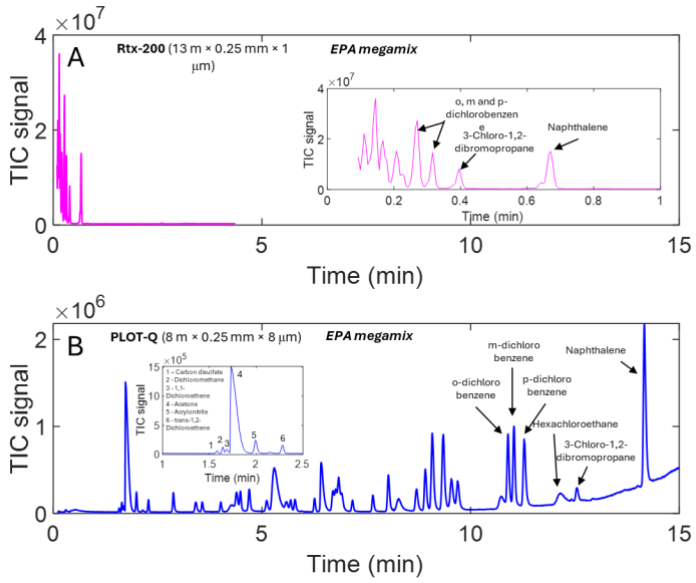


Figure 1.3. Comparison between separation of 200 ppm EPA Appendix IX Volatiles Calibration mix (Sigma-Aldrich, St.Louis, MO, USA) on (A) a wall coated open tubular column (Rtx-200) and (B) a porous layer open tubular column (PLOT-Q) ran at the same chromatographic conditions, while only changing the column. 2 mL of sample were injected both times using 1:25 split, columns were run at 1.5 ml/min carrier gas, He, flow rate and analyzed using a quadruple mass spectrometer. Temperature program was stared at 100 °C and ramped at 10 °C/min to 250 °C. Lightest compounds in the mix, that separated out using PLOT-Q column, but did not using Rtx-200 column were identified in the insert of part (B). Heaviest compounds separating on both columns were identified: o-dichlorobenzene (b.p. 180 °C), m-dichlorobenzene (b.p. 173 °C), p-dichlorobenzene (b.p. 174 °C), hexachloroethane (b.p. 187 °C), 3-chloro-1,2-dibromopropane (b.p. 196 °C) and naphthalene (b.p. 218 °C).

PLOT columns have unique selectivity and provide a great alternative to wall coated open tubular (WCOT) columns or packed columns for separations of light samples [23–25]. EPA Appendix IX volatiles calibration megamix was separated by GC using the same temperature program and chromatographic conditions on a WCOT column (Rtx-200) in Fig. 1.3. A, and on a PLOT-Q column in Fig. 1.3. B. Most of the mix eluted at the beginning of the chromatogram using a WCOT column, while using the PLOT-Q column the entire mixture was well retained and separated, everything from carbon disulfate (BP: 46 °C) to naphthalene (BP: 218 °C). In this example, even a very short length PLOT column provides unique selectivity for low boiling point analytes relative to the selectivity that is provided using a WCOT column. A heavier sample of gasoline was used to further show the unique separation abilities of PLOT columns. Gasoline is one of the heaviest “common” samples that can be run on PLOT columns, and in

Figure 1.4. we present optimized gasoline separations on a porous polymer (PLOT-S) column (Fig. 1.4. A) and on alumina oxide (Rt-Alumina BOND/MAPD) column (Fig. 1.4. B). The heaviest compound that comes elutes from both columns is 1-methyl naphthalene, which has a boiling point of 244 °C. While both columns provide intriguing separations of gasoline with relatively short run times, Rt-Alumina BOND/MAPD provides a better separation of very similar molecules, which is shown in Fig. 2C-D, where neopentane, 1-pentane, isopentane and pentane are separated into four peaks on Rt-Alumina BOND/MAPD (Fig. 1.4. D), while on PLOT-S column (Fig 1.4. C) these compounds are overlapped into two peaks. Rt-Alumina BOND/MAPD also provides a mix of sharper peaks and broader peaks. These separations show the ability to apply PLOT columns for samples containing compounds with a wide variety of boiling points, eg., isobutane (BP: -11 °C), to 1-methyl naphthalene (BP: 244 °C), which will further be discussed in Chapters 3 and 4.

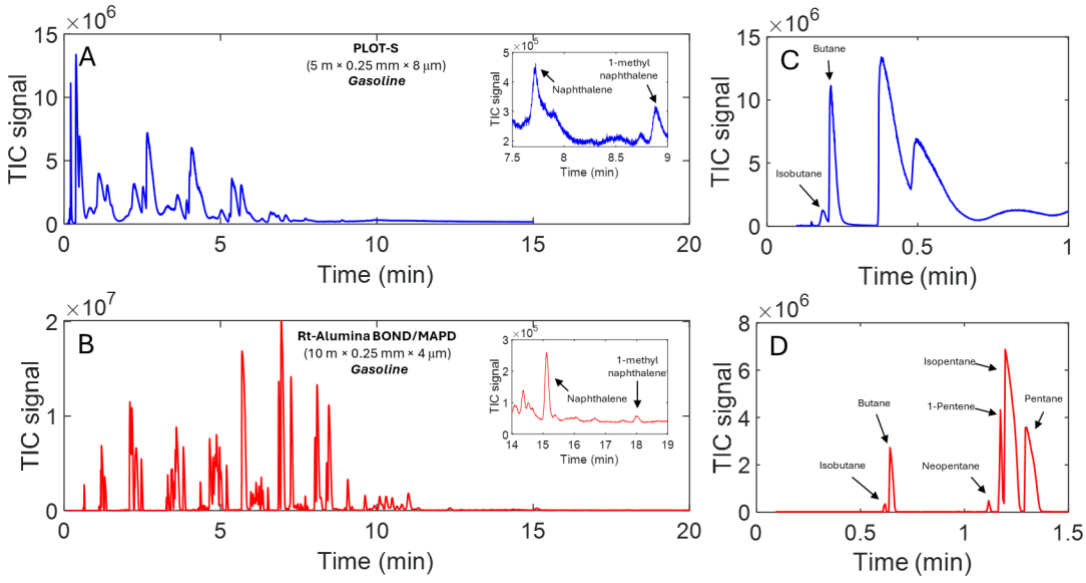


Figure 1.4. Optimized separations of gasoline that was sourced from a local gas station using two different PLOT columns. (A) Separation was performed using PLOT-S column, with 0.1 mL of sample being injected with 20:1 split, where the column was operated at 1.4 ml/min flow rate of He carrier gas and analyzed by quadruple mass spectrometer. Temperature program was started at 100 °C and ramped at 15 °C/min to 250 °C. (B) Separation was performed using Rt-Alumina BOND/MAPD column, with 0.5 mL of sample being injected with 100:1 split, where the column was operated at 1.5 ml/min flow rate of He carrier gas and analyzed by quadruple mass spectrometer. Temperature program was started at 100 °C and ramped at 20 °C/min to 250 °C. Inserts in (A) and (B) show the

heaviest compounds that are separated on these columns: naphthalene (b.p. 218 °C) and 1-methyl naphthalene (b.p. 244 °C). (C) and (D) show the beginning of gasoline separation on PLOT-S and Rt-Alumina BOND/MAPD columns, respectively. Isobutane and butane separate well at the beginning of the separation on both columns, but C5 components separate into 4 peaks on Rt-Alumina BOND/MAPD column, while they are overlapped on PLOT-S column (C).

1.2.1.2. PLOT columns in GC×GC systems

Implementation of PLOT columns has also emerged in the field of comprehensive two-dimensional (2D) gas chromatography (GC×GC), where the complementary separation mechanisms provided by two different PLOT columns can be leveraged to produce intriguing 2D separations [26,27]. In the reported cases, the samples separated are very light and relatively non-complex mixtures [26,27]. When two PLOT columns of different phases are paired to perform GC×GC separation, they are usually used in the traditional length combination for ^{1}D and ^{2}D columns, here [27] they were 30 m and 5 m respectively. Another way that PLOT columns have been used in GC×GC, is by utilizing them in the second dimension [21,28]. Due to PLOT column having very high selectivity for high volatility compounds, it is hard to pair it with a WCOT column, as there are large differences in temperatures that are needed for each of the columns to elute same compound. That in the literature is usually addressed by using a very short, for examples 8 or 14 cm, 2D PLOT column [28]. Another way that PLOT columns have been used in GC×GC is by using them as a relatively short (~0.4 m) pre-column in pair with ^{1}D column [29]. Last use of PLOT columns has inspired the GC×GC system described in Chapters 3 and 4, where PLOT column is used as a ^{1}D column, and is paired with WCOT column in the second dimension. To our knowledge, that is the first time such a GC×GC system has been used.

1.2.2 Modulator

Modulator is often called “the heart” of the GC×GC system. It reinjects small aliquots from the first dimension (^{1}D) column onto the second dimension column every couple of seconds [30]. The exact time the reinjection onto the second dimension takes is called modulation period,

or P_M. P_M is also the second dimension (2D) separation time. Typical modulation period is between 1 and 5 seconds [31–34]. A very important parameter of any modulator is its duty cycle [35]. The duty cycle represents what percentage of the analytes coming out the first dimension column get reinjected onto the second dimension column. The higher the duty cycle of the modulator, the more signal potentially can be produced by a sample, which can be important when analyzing trace compounds. Modulators come in two different types: thermal and flow modulators [30]. Thermal modulators trap and reinject analytes based on temperature, while flow modulators use gas flow to isolate the aliquots and redirect them onto the second dimension column. When it comes to modulators, both types have advantages and downsides that will be discussed in the following sections [30,35,36].

1.2.2.1 Thermal modulators

As mentioned above, thermal modulators are based on temperature and are by far the most popular type of modulator. Thermal modulators, use low temperature to trap and focus the compounds, and rapid high temperature to reinject them in a short pulse onto the second dimension. This is very reliably performed using cryogenic modulators that require a cryogen like liquid nitrogen. Recently there has been a higher number of cryogen-free modulators, which are more cost effective, as liquid nitrogen is relatively expensive, and perform temperature change electronically [30,37]. There are many commercially available cryogen-based and cryogen-free modulators available, like LECO's QUADJET Thermal modulator [38], or SepSolve's INSIGHT-Thermal modulator [39]. This makes it easier for the user to use these thermal modulators, as they have more customer support available. Further, thermal modulators typically have 100% duty cycle, meaning that they transfer everything from the first dimension to second dimension column [30].

There are also a couple of disadvantages to using a thermal modulator, main one being, that as it uses temperature to trap analytes, this type of modulator does not work well for analytes smaller than C7, which excludes a lot of lighter mixtures that could be analyzed with GC×GC [30]. Further some thermal modulators have a limitation on their minimal modulation period (minimum of 1 s), which makes it impossible to use these modulators for fast GC×GC separations [30].

1.2.2.2 Flow modulators

There are many different flow modulators that have been developed for GC×GC systems. The ones discussed here will be the dynamic pressure gradient modulator (DPGM) and high-temperature diaphragm valve, as they are relevant to work presented in Chapters 2 - 4. Generally, flow modulators use flow to trap and concentrate a section of eluent off the first dimension and redirect it onto the second dimension column. DPGM was developed in 2020 by Trinklein et al. and is a 100% duty cycle flow modulator, that if needed, can perform at very low modulation periods, of 50 ms [40]. DPGM is relatively easy to install, as it utilizes a 2-way solenoid valve and a T-union, and is a great low-cost option, especially compared to thermal modulators [40]. Due to it's high flow in the second dimension, it provides a lot of compression on the analytes and results in a large signal enhancement (from 7 to 87 times depending on analyte), but the high flow does not pertain it from being used in conjunction with time-of-flight mass spectrometer for a detector [40,41]. Another modulator that we need to discuss is the 6-port high-temperature diaphragm valve, which has a sample loop and allows for uncoupled ^{1}D and ^{2}D column flows [11,42]. First time a diaphragm valve modulator was used in 1998, but had a temperature limitation of 175 °C, which made it hard to use with higher boiling point analytes, which would need a higher GC oven temperature [43]. The high-temperature diaphragm valve can withstand

temperatures up to 325 °C, which allows for direct placement of the valve inside of the GC oven [11]. High-temperature diaphragm valve has been shown to have very low percent relative standard deviation (%RSD) for ^{2}D peak widths (≤ 3.0%), retention time (≤ 0.5%) and measured peak areas (4.4 %) [11]. The main downside of the diaphragm valve is the duty cycle that is lower than 100%, but that can be offset by having no or very little split, which is described further in Chapter 3.

1.2.3 Detector

Last, and an essential component of a GC×GC system is the detector, and there are many that are used, but the ones that are the most popular are flame ionization detectors (FID) and a mass spectrometer (MS). The most important part of the detector, when used in conjunction with GC×GC is that would have a fast enough scan rate, that second dimension peaks, that are very narrow (≥ 10 ms), would be sampled at least 7 times [40,44]. FID detectors were first introduced in 1958, and remain a popular option, due to their low cost, and high sensitivity among many others [45]. Carbon-based analytes that pass through FID flame are ionized, and the current produced by the ions (converted to voltage and processed) is measured as detector's output [46]. However, FID only provides information about how much of analyte is present in the samples but cannot be used as means of identification of the analyte. But even with that limitation there is a some work being done with GC×GC-FID systems in the recent years [5,13,47].

Another very popular detector is the mass spectrometer, which adds another dimension of data to GC×GC and allows for the identification of analytes in a mixture. Mass spectrometry works by first ionizing the compound, which then undergoes fragmentation [48]. Those fragments are then separated based on their mass-to-charge ratio in the mass spectrometer and detected based on their abundance [48]. Due to this, it has grown in popularity as a detector for

GC×GC and has a range of options from cheaper with slower scan rates (like quadrupole mass spectrometer - qMS), to more expensive ones (like high resolution mass spectrometer - HRMS) that allow for detection of analytes to the nearest 0.001 of atomic mass unit (AMU) [49]. Downside to using a MS detector, is that it cannot be used for quantification without using an internal standard [50]. Despite this, there are many recent reports of various studies performed with GC×GC-qMS [51,52], GC×GC-TOFMS [53,54] and GC×GC-HRMS [55,56]. However to address this issue with lack of quantification with the mass spectrometer, there is a growing interest in using dual detection systems, where a flow off the ^{2}D column is split between two detectors, like MS and FID, for a GC×GC-FID/MS system that allows to use FID for quantification and MS for identification [9,57,58].

1.3 Data visualization

When it comes to GC×GC data, it is very important to understand the data structure for correct analysis and processing of the chromatograms. The full visualization is shown in Figure 1.5. First (Fig. 1.5. A) we see what looks to be one peak (in blank trace), but it is made from two analytes: A (in red) and B (in blue). When GC×GC modulator transfers sections of eluent from the first dimension, it does so as often as defined by the modulation period (P_M), which in this example is 2 seconds (Fig. 1.5. B). It is important to choose a modulation period that would sample a one-dimensional peak 3-4 times (as shown in the example), this way the largest signal enhancement can be achieved by GC×GC. Which means that every 2 seconds a ^{2}D separation is performed, as seen in Fig. 1.5. B, with analyte A being less retained on ^{2}D column than analyte B. Every one of those separations then get cut and folded into a matrix, which can be seen in Fig. 1.5. C. Here, we can see ^{1}D time separation vs ^{2}D time separation vs signal of the analyte. Some researchers in the field use this visualization method, but more often, a simplified version is

presented (Fig. 1.5. D), where the signal is summed up and shown in a color scale to indicate how intense peaks are. Using this method, we can easily measure peak widths, find retention time, and interpret the chromatogram.

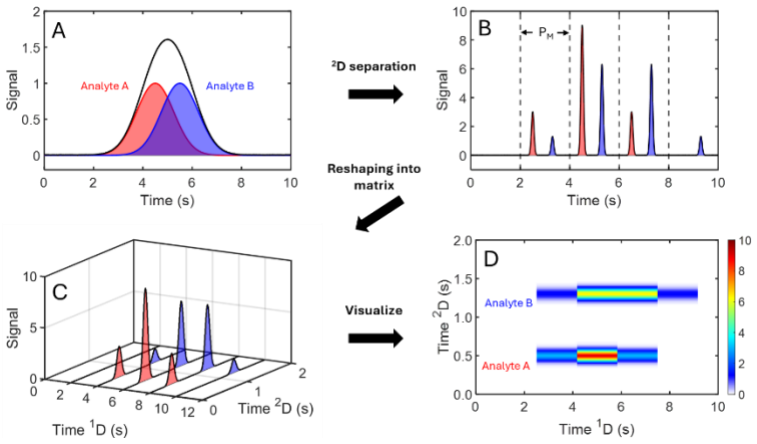


Figure 1.5. Schematic of GC×GC modulation and data visualization in two-dimensional space.

1.4 Overview of chemometric methods

Chemometrics is a field of study that strives to answer chemistry-related questions by applying mathematical and statistical methods to the data [59]. Generally, there are two types of analysis, targeted and non-targeted. Targeted analysis involves specific pre-selected analytes, their identification, quantification, and statistical significance. Non-targeted analysis is interested in every analyte that has class-distinguishing properties, that is present in the sample. Further within the non-targeted analysis, there are two types of methods: unsupervised and supervised. Unsupervised methods do not use class information when analyzing data, while supervised methods do. Chemometrics is essential to analyzing and understanding GC×GC datasets, which are data-rich and, within one typical chromatogram (30 min 1D time x 2 s 2D time x 100 Hz), can have more than 120,000 data points and over 500 analyte peaks. This requires computing power and automated analytical methods for fast and accurate analysis. Most of this work will focus specifically on analyzing biological GC×GC data with widely used methods, like F-ratio

and PCA, but also adjusting these methods and creating post-processing methods to fit the specific needs of the dataset.

1.4.1 Non-targeted Analysis Methods

Non-targeted analysis methods focus on all class-distinguishing analytes. There are many methods for non-targeted analysis, but the two that pertain most to this work are principal component analysis (PCA) and Fisher ratio. PCA is an unsupervised method that does not require assigning data classes before the analysis, while F-ratio is a supervised analysis method that does require knowing sample classes in advance.

1.4.1.1. Unsupervised Analysis Methods

PCA is one of the most used unsupervised analysis methods. It is a dimensionality reduction technique that aims to extract only the most important information from the data and represent similarity in new variables called principal components [60]. Briefly, mathematically PCA depends on the singular value decomposition (SVD) of the data matrix [60]. The main output of PCA is a scores plot, which represents how each sample projects onto lower-dimensionality variables. In a perfect scenario, similar samples will cluster next to one another, and different samples will cluster apart. Another useful output of PCA is a loadings plot, which represents how much each variable influenced each principal component [61]. PCA can be applied to GC×GC data in different ways. One way to apply it is directly to the pixel-based GC×GC data [62], another one is to apply it post-processing, like when it is applied to a peak table [63]. PCA is a very powerful technique for the classification of analytes. While it can generally cluster the analytes/samples into groups based on similarity, it is hard to define groups based solely on PCA [64].

1.4.1.2. Supervised Analysis Methods

Supervised analysis methods depend on class knowledge to identify components that cause differences in those classes. One of the most widely used methods is the Fisher ratio (F-ratio) analysis, introduced by Fisher in the 1920s [65]. This analysis method is based on analysis of variance (ANOVA) and uses the following formula [66] to calculate the difference in analyte signals between 2 classes:

$$F - ratio = \frac{Between\ Class\ Variance}{\sum(Within\ Class\ Variance)}$$

This allows for finding analytes with signals that are the most different between two classes but also with a small within-class variance. Accounting for within-class variance becomes very important when working with biological samples, where natural biological variation can be high. In these complex data sets, where variation between replicates can be high due to biological processes, finding analytes that are class-distinguishing can be very challenging. While the initial F-ratio method for GC×GC data was pixel-based [67–69], development of tile-based F-ratio software [70,71] have made it much faster to analyze GC×GC data. It can address the slight misalignment between samples and has improved sensitivity over a pixel-based method [70].

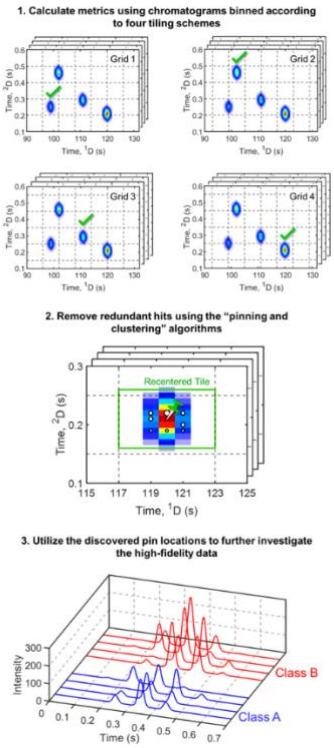


Figure 1.6. Explanation of the tile-based Fisher ratio (F-ratio), four tiling schemes, and "pinning and clustering".

The way the F-ratio tile method works is shown in Figure 1.6. First, the user needs to select a tile size, which, in general, needs to be slightly bigger than the peak size and account for any peak retention time shifting between chromatograms. Trinklein et al. have published guidelines on what tile size might be best for specific scenarios [72]. However, it still is a parameter that every user has to optimize based on their data. Further, a single grid cannot be used to perform tile-based F-ratio analysis, due to potential peak splitting, so a four-grid scheme (Fig.1.6.) is used instead, and that way the best grid (the one that centers the most around the specific peak) is automatically chosen for that analyte. At this point F-ratio metric is calculated for every hit, based on signal-to-noise and other parameters set by the user. While the four-grid system is great for accurate peak sampling, it also produces redundant sampling of each hit, called "redundant hits". These are then removed through a "pin and clustering" step of the software, by finding the peak maxima within a tile, and all the hits within a user defined window

called “cluster window” are conglomerated into one hit. From the final pin location, tile gets recentered around the peak and the 2D data can be reversed to vector form if needed for further analysis (Fig.1.6.).

The F-ratio tile method has been commercialized through LECO’s ChromaTOF Tile software [73]. This has made it easier than ever to analyze GC×GC data for both expert and beginner users, which is exciting due to GC×GC being more of the technique of academic environment than industry due to the lack of highly trained workers with expertise in GC×GC instrumentation and data analysis, that could get through a high threshold of entry to use this technique in more manufacturing settings. There has been a growing number of reports where ChromaTOF Tile software has been used to analyze samples in food analysis [74,75], environmental analysis [76] and other areas [77].

1.4.2. Post-processing of F-ratio Results

After F-ratio results are obtained there sometimes is a need to post-process the data. This usually is due to complex analytical questions that have to be answered by the data, like looking for analytes that not only have differences between classes but also express a specific pattern over time (Chapter 5), there are more than one change that is happening within the data (Chapter 6), or we need to find relationships between classes of the data (Chapter 7). All of these post-processing can always be done manually, however, when looking at hundreds of F-ratio hits, it is always better to automate these decisions by using a specific metric [78] or by fitting simulated data to find patterns [79]. Another post-processing that always must be addressed in the data is finding cut-off threshold for F-ratio, that would capture all the analytes that are significant to the analysis without introducing false negatives [71,80].Within this work, different methods for finding thresholds will be presented including simulating random data based on the real data

(Chapter 5) and creating pseudo-null distributions using randomized signals of real data (Chapter 6).

1.5. Overview of the following chapters

1.5.1 Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection

A baseline correction method is developed for comprehensive two-dimensional (2D) chromatography (GC×GC) with flame-ionization detection (FID) using dynamic pressure gradient modulation (DPGM). The DPGM-GC×GC-FID utilized porous layer open tubular (PLOT) columns in both dimensions to focus on light hydrocarbon separations. Since DPGM is nominally a stop-flow modulation technique, a rhythmic baseline disturbance is observed in the FID signal that cycles with the modulation period (P_M). This baseline disturbance needs to be corrected to optimize trace analysis. The baseline correction method has three steps: collection of a background "blank" chromatogram and multiplying it by an optimized normalization factor, subtraction of the normalization-optimized background chromatogram from a sample chromatogram, and application of Savitzky-Golay smoothing. An alkane standard solution, containing pentane, hexane and heptane was used for method development, producing linear calibration curves ($r^2 > 0.991$) over a broad concentration range (7.8 ppm – 4000 ppm). Further, the limit-of-detection (*LOD*) and limit-of-quantification (*LOQ*) were determined for pentane (*LOD* = 2.5 ppm, *LOQ* = 8.2 ppm), hexane (*LOD* = 0.9 ppm, *LOQ* = 3.0 ppm), and heptane (*LOD* = 1.9 ppm, *LOQ* = 6.4 ppm). A natural gas sample separation illustrated method applicability, whereby the DPGM produced a signal enhancement (*SE*) of 30 for isopentane, where *SE* is defined as the height of the tallest ^{2}D peak in the modulated chromatogram for the analyte divided by the height of the unmodulated ^{1}D peak. The 30-fold *SE* resulted in about a

10-fold improvement in the signal-to-noise ratio (*S/N*) for isopentane. Additional versatility of the baseline correction method for more complicated samples was demonstrated for an unleaded gasoline sample, which enabled the detection (and visual appearance) of trace components.

1.5.2 Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension

A comprehensive two-dimensional gas chromatography instrument with quadrupole mass spectrometry (GC×GC-qMS) is developed and evaluated, by retrofitting a commercial GC-qMS instrument with a high temperature diaphragm valve (HT-DV) modulator. A porous layer open tubular (PLOT) column in the first dimension (^{1}D), and a more polar wall coated open tubular (WCOT) column in the second dimension (^{2}D) were used. The HT-DV can modulate analytes of any volatility and the combination of the PLOT and WCOT columns provides ample separation selectivity for samples containing an extended range of volatility from highly volatile analytes to analytes of more moderate volatility. In addition to instrument development, we describe data visualization considerations when upgrading from GC-qMS to GC×GC-qMS. Unlike time-of-flight mass spectrometry (TOFMS), the mass spectrum cycle time for qMS (defined as the sum of scan time and interscan delay) is a linear function of the range of mass channels (*m/z*) analyzed. If the cycle time does not divide evenly into the modulation period, P_M, visual distortions will appear in the GC×GC chromatogram. We present two approaches to address this GC×GC chromatogram distortion issue. The first method does not require post-run computations; however, care must be taken in selecting an appropriate mass range to minimize the visual distortion issue. A half of the modulation period of total shift in the ^{2}D separations with a P_M of 2 s was determined tolerable for a 15 min separation, with 44 mass ranges (and their accompanying scan rate) deemed suitable out of 350 possible mass ranges. Also, a computational

"resampling" method, which can be used to correct any qMS scan rate used was implemented. Figures-of-merit for the HT-DV modulated GC×GC-qMS instrument were also examined. Reproducible peak height (average RSD_{height} = 9.2 %, s = 3.5 %) and areas (average RSD_{area} = 9.3 %, s = 2.8 %) were obtained with low *LOD*s (average *LOD* = 4.2 mg/L, s = 2.5 mg/L).

1.5.3 Performance of thin-film porous layer open tubular columns in one-dimensional and comprehensive two-dimensional gas chromatography-mass spectrometry systems

The performance evaluation of thin-film (4 µm and 2 µm) porous layer open tubular (PLOT) columns compared to commercially available 8 µm film thickness PLOT-S columns. Currently, there is limited research done exploring the possibility of thinner-film PLOT columns, instead most of the research is in column manufacturing and new phases. Further, currently, these columns are more used in liquid than gas chromatography. Finally PLOT columns have been very underutilized in GCGC due to their extreme separation capability and being hard to balance with the WCOT columns to get perpendicular separations. Here we first compare one-dimensional separations between 8 µm film thickness PLOT-S columns and 4 µm and 2 µm film thickness columns and found that the column with the thinnest film has a higher boiling point range by 55 °C compared to the 8 µm film PLOT-S column. Further, through separations of the refinery mix 2 µm film column is shown not to lose its selectivity for extremely light analytes. This further was shown by comparing a full separation of the Refinery mix on a 2 µm film PLOT-S column with the one performed on an Rxi-5 wall-coated column. We also explored some of the GC×GC separations comparing 8 µm and 2 µm film PLOT-S columns used in the first dimension and their performance separating Calibration MegaMix, Refinery Mix, and a real-world sample of gasoline. These separations on a thinner-film column showed that the

column utilizes the separation space better compared to commercially available columns, by not losing separation on the light components, but separating them at lower temperatures.

1.5.4 Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry

A computational method for the untargeted determination of cycling yeast metabolites using a comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) dataset is presented. The yeast metabolomic cycle for the diploid yeast strain CEN.PK with a 5 hr cycle period relative to the O_2 concentration level is comprehensively examined to determine the metabolites that exhibit cycling. Samples were collected over only two cycles (10 hr with a total of 24 time-point sampling intervals at 25 min each) as an experimental constraint. Due to the limited number of cycles expressed in the dataset, a computational method was devised to determine with statistical significance whether or not a given metabolite exhibited a temporal signal pattern that constituted cycling in the context of the 5 hr cycle period. The computational method we report compares the experimentally obtained 24 time-point metabolite signal sequences to randomly generated signal sequences coupled with statistically based confidence level *LOF* metrics to determine whether or not a given metabolite expresses cycling, and if so, what is the phase of the cycling. Initially the GC×GC-TOFMS dataset was analyzed using tile-based Fisher ratio (F-ratio) analysis. Since there were 24 time-point intervals, this constituted 24 sample classes in the F-ratio calculation which produced 672 metabolite hits. Next, application of the computational method determined that there were 210 of the 672 metabolites exhibiting cycling: 55 identified metabolites and 155 unknown metabolites. Furthermore, the 210 cycling metabolites were categorized into four groups, and where

applicable, a phase determined: 1 cycle/5 hr period (106 metabolites), 2 cycles/5 hr period (13 metabolites), spiky pattern (12 metabolites), or multimodal pattern (79 metabolites).

1.5.5 Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data

Herein, two "orthogonal" characteristics of moisture damaged cacao beans (temporally dependent molding kinetics versus the time-independent geographical region of origin) are simultaneously analyzed in a comprehensive two-dimensional (2D) gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) dataset using tile-based Fisher ratio (F-ratio) analysis. Cacao beans from six geographical regions were analyzed once a day for six days following the initiation of moisture damage to trigger the molding process. Thus, there are two "extremes" to the experimental sample class design: six time points for the molding kinetics versus the six geographical regions of origin, resulting in a 6 × 6 element signal array referred to as a *composite chemical fingerprint* (CCF) for each analyte. Usually, this study would involve initial generation of two separate hit lists using F-ratio analysis, one hit list from inputting the data with the six time point classes, then another hit list from inputting the dataset from the perspective of geographic region of origin. However, analysis of two separate hit lists with the intent to distill them down to one hit list is extremely time-consuming and fraught with shortcomings due to the challenges associated with attempting to match analytes across two hit lists. To address this challenge, tile-based F-ratio analysis is "orthogonally applied" to each analyte CCF to simultaneously determine two F-ratios at the chromatographic 2D location (F-ratio$_{kinetic}$ and F-ratio$_{region}$) for each hit, by ranking a single hit list using the higher of the two F-ratios resulting in the discovery of 591 analytes. Further, using a pseudo-null distribution

approach, at the 99.9 % threshold over 400 analytes were deemed suitable for PCA classification. Using a more stringent 99.999 % threshold, over 100 analytes were explored more deeply using PARAFAC to provide a purified mass spectrum.

1.5.6 Tile-based analysis with support vector machine regression modeling of GC×GC-TOFMS data of VOCs produced by Malassezia pachydermatis grown at variable pHs

Malassezia yeasts are a commensal microorganism found in human and animal skin. Species of *Malassezia* have been connected to skin and opportunistic infections where certain microenvironmental conditions are required in the host for the pathogenic processes to occur. We present the analysis of the volatile space of *Malassezia pachydermatis* at three pHs (5.7, 9.7, and 12.4) by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS). Since changes in pH also affect the growth media and the VOCs produced by it, media blanks at the three pHs were analyzed, with 5 replicates of each of the 6 samples. Following data collection, GC×GC-TOFMS chromatograms were analyzed by Fisher ratio software that found 566 analytes, out of which 288 were tentatively identified with a mass spectrum match value (MV) ≥ 800 based upon a NIST library search. A signal pattern for each of the 566 analytes was obtained by averaging the replicates, and two metrics (*R* and *RSD*) were calculated for each signal pattern. The *R metric* was defined to focus upon the differences between analyte signals of blanks and *M. pachydermatis* by taking away the influence of pH changes. The *RSD* metric, on the other hand, was defined to evaluate only the influence of pH. Based on the *R* metric magnitude, the analytes were split into 3 categories: media analytes consumed (in media only), pH-dependent analytes (in media and *M. pachydermatis*), and analytes produced in *M. pachydermatis* only. Many of the *M. pachydermatis* produced analytes were already shown to be produced by other yeast species and shown to have biological

significance. Further, there is evidence of some bioconversions between the consumed analytes discovered versus the analytes produced. We also verified our classification results using the support vector machines (SVM) model, where cross-validation showed very promising results with TPR and TNR both being over 0.95 and error being below 0.03 (or 3%).

1.6 References

[1] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, Journal of Chromatographic Science 29 (1991) 227–231. https://doi.org/10.1093/chromsci/29.6.227.

[2] J.B. de Oliveira, M.F.M. Carvalho, P.P. de Souza, Z. de Lourdes Cardeal, Development and Application of Novel Extraction Method to Analyze Herbal Supplements and Adulterant Determination by GCxGC/Q-TOFMS, Food Anal. Methods (2024). https://doi.org/10.1007/s12161-024-02657-y.

[3] T.J. Trinklein, J. Jiang, R.E. Synovec, Profiling Olefins in Gasoline by Bromination Using GC×GC-TOFMS Followed by Discovery-Based Comparative Analysis, Anal. Chem. 94 (2022) 9407–9414. https://doi.org/10.1021/acs.analchem.2c01549.

[4] C.N. Cain, G.S. Ochoa, R.E. Synovec, Enhancing partial least squares modeling of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data by tile-based variance ranking, Journal of Chromatography A 1694 (2023) 463920. https://doi.org/10.1016/j.chroma.2023.463920.

[5] K.E. Ikonen, W. Li, A. Ruonakangas, G. Kilaz, H.I. Kenttämaa, Different Hydrocarbon Types in Jet Fuel as Determined by GCxGC/FID - An Alternative Method to D2425, Energy Fuels 37 (2023) 6629–6641. https://doi.org/10.1021/acs.energyfuels.3c00774.

[6] G.S. Ochoa, R.E. Synovec, Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, Talanta 259 (2023) 124525. https://doi.org/10.1016/j.talanta.2023.124525.

[7] P.C.K. Makhubela, E.R. Rohwer, Y. Naudé, Detection of tuberculosis-associated compounds from human skin by GCxGC-TOFMS, Journal of Chromatography B 1231 (2023) 123937. https://doi.org/10.1016/j.jchromb.2023.123937.

[8] E. Barberis, E. Amede, S. Khoso, L. Castello, P.P. Sainaghi, M. Bellan, P.E. Balbo, G. Patti, D. Brustia, M. Giordano, R. Rolla, A. Chiocchetti, G. Romani, M. Manfredi, R. Vaschetto, Metabolomics Diagnosis of COVID-19 from Exhaled Breath Condensate, Metabolites 11 (2021) 847. https://doi.org/10.3390/metabo11120847.

[9] J.M. Byrne, L.M. Dubois, J.D. Baker, J.-F. Focant, K.A. Perrault, A non-targeted data processing workflow for volatile organic compound data acquired using comprehensive two-dimensional gas chromatography with dual channel detection, MethodsX 7 (2020) 101009. https://doi.org/10.1016/j.mex.2020.101009.

[10] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, Journal of Chromatography A 1383 (2015) 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.

[11] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, Journal of Chromatography A 1424 (2015) 127–133. https://doi.org/10.1016/j.chroma.2015.10.098.
[12] C.-F. Ou-Yang, Y.-J. Chen, H.-C. Hsieh, C.-T. Lee, K.-H. Chi, N.-H. Lin, C.-C. Chang, J.-L. Wang, Identification of organic constituents on atmospheric particulate matter in the East Asian background air of free troposphere by GC×GC-TOFMS, Chemosphere 364 (2024) 143095. https://doi.org/10.1016/j.chemosphere.2024.143095.
[13] M. Biedermann, K. Grob, Advantages of comprehensive two-dimensional gas chromatography for comprehensive analysis of potential migrants from food contact materials, Analytica Chimica Acta 1057 (2019) 11–17. https://doi.org/10.1016/j.aca.2018.10.046.
[14] Restek's PLOT Column Family, (2020). https://www.restek.com/en/technical-literature-library/articles/resteks-PLOT-column-family/ (accessed July 20, 2022).
[15] J.J. Kirkland, Fibrillar Boehmite-A New Adsorbent for Gas Solid Chromatography., Anal. Chem. 35 (1963) 1295–1297. https://doi.org/10.1021/ac60202a038.
[16] D.L. Petitjean, C.J. Leftault, Oxide-Coated Aluminum Tubing for Capillary Gas Chromatography, Journal of Chromatographic Science 1 (1963) 18–21. https://doi.org/10.1093/chromsci/1.3.18.
[17] E. de Vanssay, P. Capilla, D. Coscia, L. Do, R. Sternberg, F. Raulin, Gas chromatography of Titan's atmosphere: IV. Analysis of permanent gases in the presence of hydrocarbons and nitriles with a Molsieve PLOT capillary column, Journal of Chromatography A 639 (1993) 255–259. https://doi.org/10.1016/0021-9673(93)80261-6.
[18] H. Yun, M.L. Lee, Fast gas chromatography of light hydrocarbons and permanent gases on porous-layer open-tubular columns, Field Analytical Chemistry & Technology 1 (1996) 60–64. https://doi.org/10.1002/(SICI)1520-6521(1996)1:1<60::AID-FACT8>3.0.CO;2-I.
[19] H. Yun, M.L. Lee, K.E. Markides, Charcoal porous layer open tubular column gas chromatography for permanent gas analysis, Journal of Microcolumn Separations 7 (1995) 207–212. https://doi.org/10.1002/mcs.1220070304.
[20] Z. Ji, I.L. Chang, A new look at light hydrocarbon separations on commercial alumina PLOT columns: Column selectivity and separation, Journal of High Resolution Chromatography 19 (1996) 32–36. https://doi.org/10.1002/jhrc.1240190106.
[21] W.-C. Liao, C.-F. Ou-Yang, C.-H. Wang, C.-C. Chang, J.-L. Wang, Two-dimensional gas chromatographic analysis of ambient light hydrocarbons, Journal of Chromatography A 1294 (2013) 122–129. https://doi.org/10.1016/j.chroma.2013.04.008.
[22] G.L. Reid III, D.W. Armstrong, Cyclodextrin PLOT columns for the gas-solid chromatographic separation of light hydrocarbons and inorganic gases, Journal of Microcolumn Separations 6 (1994) 151–157. https://doi.org/10.1002/mcs.1220060208.
[23] C.-H. Wang, C.-Y. Huang, H.-K. Yak, H.-C. Hsieh, J.-L. Wang, Identifying an unknown compound in flue gas of semiconductor industry – Forensics of a perfluorocarbon, Chemosphere 264 (2021) 128504. https://doi.org/10.1016/j.chemosphere.2020.128504.
[24] P. Kotnik, M. Čolnik, M. Finšgar, Ž. Knez, M. Škerget, Determination of C1C6 hydrocarbons in gaseous plastic degradation products by GC–MS method, Polymer Degradation and Stability 182 (2020) 109386. https://doi.org/10.1016/j.polymdegradstab.2020.109386.
[25] E.Y. Yakovleva, Y.V. Patrushev, Determination of Hydrocarbons in n-Butane on Porous Layer Capillary Columns with Nonpolar Stationary Phase, Chromatographia 84 (2021) 1095–1104. https://doi.org/10.1007/s10337-021-04092-1.

[26] A.Y. Sholokhova, Y.V. Patrushev, V.N. Sidelnikov, A.K. Buryak, Analysis of light components in pyrolysis products by comprehensive two-dimensional gas chromatography with PLOT columns, Talanta 209 (2020) 120448. https://doi.org/10.1016/j.talanta.2019.120448.
[27] Y.V. Patrushev, V.N. Sidelnikov, Selection of the porous layer open tubular columns for separation of light components in comprehensive two-dimensional gas chromatography, Journal of Chromatography A 1579 (2018) 83–88. https://doi.org/10.1016/j.chroma.2018.10.015.
[28] D. Peroni, R.J. Vonk, W. van Egmond, H.-G. Janssen, Macroporous polymer monoliths as second dimension columns in comprehensive two-dimensional gas chromatography: A feasibility study, Journal of Chromatography A 1268 (2012) 139–149. https://doi.org/10.1016/j.chroma.2012.10.019.
[29] X. Guan, Z. Zhao, S. Cai, S. Wang, H. Lu, Analysis of volatile organic compounds using cryogen-free thermal modulation based comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry, Journal of Chromatography A 1587 (2019) 227–238. https://doi.org/10.1016/j.chroma.2018.12.025.
[30] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, TrAC Trends in Analytical Chemistry 113 (2019) 379–391. https://doi.org/10.1016/j.trac.2018.04.016.
[31] D.J. Oliva, C. Cheung, K.A. Perrault, Fisher ratio feature selection by manual peak area calculations on comprehensive two-dimensional gas chromatography data using standard mixtures with variable composition, storage, and interferences, Anal Bioanal Chem 415 (2023) 2575–2585. https://doi.org/10.1007/s00216-022-04484-8.
[32] Y. Zou, P.-H. Stefanuto, M. Maimone, M. Janssen, J.-F. Focant, Unraveling the Complex Olefin Isomer Mixture Using Two-Dimensional Gas Chromatography-Photoionization-Time of Flight Mass Spectrometry, Journal of Chromatography A 1645 (2021) 462103. https://doi.org/10.1016/j.chroma.2021.462103.
[33] N. Sun, P. Krishnan, C.A. Rees, M. Zhang, K.A.J.M. Stevenson, J.E. Hill, Profiling volatile organic compounds from human plasma using GC × GC-ToFMS, J. Breath Res. 17 (2023) 037104. https://doi.org/10.1088/1752-7163/acd806.
[34] E.A.H. Keppler, M.C.C.V. Dyke, H.L. Mead, D.F. Lake, D.M. Magee, B.M. Barker, H.D. Bean, Volatile metabolites in lavage fluid are correlated to Valley fever disease severity in murine model lung infections, (2022) 2022.10.02.510537. https://doi.org/10.1101/2022.10.02.510537.
[35] P.Q. Tranchida, G. Purcaro, P. Dugo, L. Mondello, G. Purcaro, Modulators for comprehensive two-dimensional gas chromatography, TrAC Trends in Analytical Chemistry 30 (2011) 1437–1461. https://doi.org/10.1016/j.trac.2011.06.010.
[36] M. Edwards, A. Mostafa, T. Górecki, Modulation in comprehensive two-dimensional gas chromatography: 20 years of innovation, Anal Bioanal Chem 401 (2011) 2335–2349. https://doi.org/10.1007/s00216-011-5100-6.
[37] H. Boswell, K. Tarazona Carrillo, T. Górecki, Evaluation of the Performance of Cryogen-Free Thermal Modulation-Based Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GC×GC-TOFMS) for the Qualitative Analysis of a Complex Bitumen Sample, Separations 7 (2020) 13. https://doi.org/10.3390/separations7010013.
[38] Thermal Modulator, LECO Corporation (n.d.). https://www.leco.com/products/thermal-modulator/ (accessed September 3, 2024).

[39] W. designed and developed by Z.- https://www.zarr.com, INSIGHT-Thermal modulator of GCxGC, (n.d.). https://www.sepsolve.com/gas-chromatography/insight-thermal-modulator-of-gcxgc.aspx (accessed September 3, 2024).
[40] T.J. Trinklein, D.V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, Journal of Chromatography A 1609 (2020) 460488. https://doi.org/10.1016/j.chroma.2019.460488.
[41] S. Schöneich, D.V. Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, Journal of Chromatography A 1620 (2020) 460982. https://doi.org/10.1016/j.chroma.2020.460982.
[42] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.
[43] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. https://doi.org/10.1021/ac980164m.
[44] M. Adahchour, M. Brandt, H.-U. Baier, R.J.J. Vreuls, A.M. Batenburg, U.A.Th. Brinkman, Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: principles and applications, Journal of Chromatography A 1067 (2005) 245–254. https://doi.org/10.1016/j.chroma.2004.09.094.
[45] T. Holm, Aspects of the mechanism of the flame ionization detector, Journal of Chromatography A 842 (1999) 221–227. https://doi.org/10.1016/S0021-9673(98)00706-7.
[46] J.V. Hinshaw, Flame Ionization: GC's Workhorse Detector, 33 (2015) 470–477. https://www.chromatographyonline.com/view/flame-ionization-gcs-workhorse-detector (accessed September 3, 2024).
[47] M. Auersvald, M. Šiman, P. Vozka, P. Straka, Quantitative determination of olefins in pyrolysis oils from waste plastics and tires using selective adsorption by Ag-SiO2 followed by GC×GC-FID, Talanta (2024) 126792. https://doi.org/10.1016/j.talanta.2024.126792.
[48] E. de Hoffmann, V. Stroobant, Mass Spectrometry: Principles and Applications, John Wiley & Sons, 2007.
[49] J.C. Cook-Botelho, L.M. Bachmann, D. French, Chapter 10 - Steroid hormones, in: H. Nair, W. Clarke (Eds.), Mass Spectrometry for the Clinical Laboratory, Academic Press, San Diego, 2017: pp. 205–230. https://doi.org/10.1016/B978-0-12-800871-3.00010-9.
[50] J. Lill, Proteomic tools for quantitation by mass spectrometry, Mass Spectrometry Reviews 22 (2003) 182–194. https://doi.org/10.1002/mas.10048.
[51] Y. Zhang, H. Ren, X. Tang, Q. Liu, W. Xiao, Z. Zhang, Y. Tian, A GC×GC-MS method based on solid-state modulator for non-targeted metabolomics: Comparison with traditional GC-MS method, Journal of Pharmaceutical and Biomedical Analysis 243 (2024) 116068. https://doi.org/10.1016/j.jpba.2024.116068.
[52] X. Jia, P. Yu, Q. An, J. Ren, G. Fan, Z. Wei, X. Li, S. Pan, Identification of glucosinolates and volatile odor compounds in microwaved radish (Raphanus sativus L.) seeds and the corresponding oils by UPLC-IMS-QTOF-MS and GC × GC-qMS analysis, Food Research International 169 (2023) 112873. https://doi.org/10.1016/j.foodres.2023.112873.
[53] C.L. Jenkins, H.D. Bean, Dependence of the Staphylococcal Volatilome Composition on Microbial Nutrition, Metabolites 10 (2020) 347. https://doi.org/10.3390/metabo10090347.

[54] N. Nouri, N. Sun, J.E. Hill, A feasibility study of sample re-collection in the analysis of selected volatile compounds in breath samples using GC×GC-TOFMS, Journal of Chromatography A 1730 (2024) 465125. https://doi.org/10.1016/j.chroma.2024.465125.
[55] R.K. Nelson, J. Forsythe, C. Eiserbeck, A.G. Scarlett, K. Grice, O.C. Mullins, C.M. Reddy, GC×GC Analysis of Novel 2α-Methyl Biomarker Compounds from a Large Middle East Oilfield, Energy Fuels 36 (2022) 8853–8865. https://doi.org/10.1021/acs.energyfuels.2c00949.
[56] P. Mara, R.K. Nelson, C.M. Reddy, A. Teske, V.P. Edgcomb, Sterane and hopane biomarkers capture microbial transformations of complex hydrocarbons in young hydrothermal Guaymas Basin sediments, Commun Earth Environ 3 (2022) 1–14. https://doi.org/10.1038/s43247-022-00582-8.
[57] J. Melder, J. Zinsmeister, T. Grein, S. Jürgens, M. Köhler, P. Oßwald, Comprehensive Two-Dimensional Gas Chromatography: A Universal Method for Composition-Based Prediction of Emission Characteristics of Complex Fuels, Energy Fuels 37 (2023) 4580–4595. https://doi.org/10.1021/acs.energyfuels.2c04270.
[58] L. Brito, F. Payan, F. Albrieux, E. Guillon, J.A. Martens, G.D. Pirngruber, Hydrocracking of a Long Chain Alkyl-Cycloalkane: Role of Porosity and Metal-Acid Balance, ChemCatChem 15 (2023) e202201286. https://doi.org/10.1002/cctc.202201286.
[59] S.D. Brown, T.B. Blank, S.T. Sum, L.G. Weyer, Chemometrics, Anal. Chem. 66 (1994) 315–359. https://doi.org/10.1021/ac00084a014.
[60] H. Abdi, L.J. Williams, Principal component analysis, WIREs Computational Statistics 2 (2010) 433–459. https://doi.org/10.1002/wics.101.
[61] T.J. Trinklein, C.N. Cain, G.S. Ochoa, S. Schöneich, L. Mikaliunaite, R.E. Synovec, Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis, Anal. Chem. 95 (2023) 264–286. https://doi.org/10.1021/acs.analchem.2c04235.
[62] J.S. Lübeck, G.L. Alexandrino, J.H. Christensen, GC × GC–HRMS nontarget fingerprinting of organic micropollutants in urban freshwater sediments, Environmental Sciences Europe 32 (2020) 78. https://doi.org/10.1186/s12302-020-00353-2.
[63] J. Dekeirsschieter, P.-H. Stefanuto, C. Brasseur, E. Haubruge, J.-F. Focant, Enhanced Characterization of the Smell of Death by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry (GCxGC-TOFMS), PLOS ONE 7 (2012) e39005. https://doi.org/10.1371/journal.pone.0039005.
[64] C. West, Statistics for Analysts Who Hate Statistics, Part III: Principal Component Analysis, 34 (2016) 868–869. https://www.chromatographyonline.com/view/statistics-analysts-who-hate-statistics-part-iii-principal-component-analysis (accessed September 4, 2024).
[65] R.A. Fisher, Statistical Methods for Research Workers, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer, New York, NY, 1992: pp. 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6.
[66] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher Ratio Method Applied to Third-Order Separation Data To Identify Significant Chemical Components of Metabolite Extracts, Anal. Chem. 78 (2006) 5068–5075. https://doi.org/10.1021/ac0602625.
[67] R. E. Mohler, K. M. Dombek, J. C. Hoggard, K. M. Pierce, E. T. Young, R. E. Synovec, Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: combining discovery-mode and deconvolution chemometric software, Analyst 132 (2007) 756–767. https://doi.org/10.1039/B700061H.

[68] E.M. Humston, K.M. Dombek, B.P. Tu, E.T. Young, R.E. Synovec, Toward a global analysis of metabolites in regulatory mutants of yeast, Anal Bioanal Chem 401 (2011) 2387–2402. https://doi.org/10.1007/s00216-011-4800-2.
[69] A.C. Beckstrom, E.M. Humston, L.R. Snyder, R.E. Synovec, S.E. Juul, Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model, Journal of Chromatography A 1218 (2011) 1899–1906. https://doi.org/10.1016/j.chroma.2011.01.086.
[70] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.
[71] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. https://doi.org/10.1021/ac504472s.
[72] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, Journal of Chromatography A 1677 (2022) 463321. https://doi.org/10.1016/j.chroma.2022.463321.
[73] LECO Corporation, ChromaTOF Tile Analytical Software, (n.d.). https://www.leco.com/product/chromatof-tile.
[74] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, Journal of Chromatography A 1662 (2022) 462735. https://doi.org/10.1016/j.chroma.2021.462735.
[75] N. Koljančić, A.A. Gomes, I. Špánik, A non-target geographical origin screening of botrytized wines through comprehensive two-dimensional gas chromatography coupled with high-resolution mass spectrometry, Journal of Separation Science 46 (2023) 2300249. https://doi.org/10.1002/jssc.202300249.
[76] D.M. Mazur, A.A. Sosnova, T.B. Latkin, B.V. Artaev, K. Siek, D.A. Koluntaev, A.T. Lebedev, Application of clusterization algorithms for analysis of semivolatile pollutants in Arkhangelsk snow, Anal Bioanal Chem 415 (2023) 2587–2599. https://doi.org/10.1007/s00216-022-04390-z.
[77] N. Koljančić, O. Vyviurska, I. Špánik, Aroma Compounds in Essential Oils: Analyzing Chemical Composition Using Two-Dimensional Gas Chromatography–High Resolution Time-of-Flight Mass Spectrometry Combined with Chemometrics, Plants 12 (2023) 2362. https://doi.org/10.3390/plants12122362.
[78] C.N. Cain, T.J. Trinklein, G.S. Ochoa, R.E. Synovec, Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification, Anal. Chem. 94 (2022) 5658–5666. https://doi.org/10.1021/acs.analchem.2c00223.
[79] T. Kasukawa, M. Sugimoto, A. Hida, Y. Minami, M. Mori, S. Honma, K. Honma, K. Mishima, T. Soga, H.R. Ueda, Human blood metabolite timetable indicates internal body time, PNAS 109 (2012) 15036–15041. https://doi.org/10.1073/pnas.1207768109.

[80] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.E. Synovec, Statistical inference of mass channel purity from Fisher ratio analysis using comprehensive two-dimensional gas chromatography with time of flight mass spectrometry data, Journal of Chromatography A 1627 (2020) 461401. https://doi.org/10.1016/j.chroma.2020.461401.

Chapter 2. Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection.

This chapter is reproduced from Lina Mikaliunaite, Paige E. Sudol, Caitlin N. Cain, Robert E. Synovec, "Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection" *J. Chromatogr. A* 1652 (2021) 462358.

2.1 Introduction

Low molecular weight and low boiling point compounds, commonly referred to as volatile organic compounds (VOCs), are critical components in a variety of samples of analytical interest, such as wastewater [1, 2], soil [3, 4], natural gas [5], and gasoline [6–8]. Since these compounds can often be present at trace levels (i.e., low ppm range), they require analysis via sensitive and robust analytical instrumentation. Natural gas is a particularly challenging sample matrix, as it contains primarily methane as well as longer and much less volatile hydrocarbons ranging from C6 to C12 [9–12].

Gas chromatography (GC) is the prominent technique for the detection and quantification of VOCs due to the high selectivity and resolution provided [1, 2, 4]. Method development for GC analysis of VOCs primarily relies upon the appropriate selection of the separation column. Primarily, GC separations are performed using wall-coated open tubular (WCOT) capillary columns, where a film of a liquid stationary phase is directly coated on the inside of the capillary. WCOT columns provide more efficient separations compared to packed columns [13]. However, very low molecular weight VOCs can exhibit poor retention on packed columns [14,15] or WCOT columns [16] due to the extreme volatility which causes them not to be baseline separated. In contrast, porous layer open tubular (PLOT) columns can be used for better separation of highly volatile, low molecular weight VOCs, as they provide unique selectivity and suitable retention for

these compounds. PLOT columns utilize a layer of a porous adsorbent like fused silica, alumina, or molecular sieves to coat the inner wall of the capillary column [13]. For example, Burger et al. recently developed a cryo-wafer device containing multiple PLOT capillary columns for headspace sampling of the C6+ fraction of natural gas, followed by desorption into a GC-MS for identification and relative quantification [17]. While PLOT columns improve the retention of these low molecular weight VOCs, coelution of isomers and a low signal-to-noise ratio (*S/N*) can still complicate the analysis.

However, the use of comprehensive two-dimensional (2D) gas chromatography (GC×GC) can overcome these challenges by providing an orthogonal separation dimension and improving signal enhancement [18–20]. GC×GC was introduced by Liu and Philips in 1991 [21], and has evolved ever since with wide applicability in many different fields like forensics [22], cosmetics [23, 24] environmental science [25, 26], and food analysis [27]. In GC×GC, the first-dimension (^{1}D) column is coupled to a second-dimension (^{2}D) column by a modulator, which nominally isolates and reinjects the ^{1}D eluate onto the ^{2}D column [28, 29]. GC×GC can provide about a 10-fold increase in peak capacity relative to one-dimensional GC [30], as well as considerable signal enhancement. Generally, thermal modulators provide the greatest signal enhancement in GC×GC work as a result of the temperature differential utilized to trap and reinject analytes, which produces a 100% duty cycle (i.e., all of the ^{1}D eluate is transferred to the ^{2}D column) [31–34]. Thermal modulators have been used in conjunction with PLOT columns in GC×GC, and although they are widely commercially available, some are expensive to operate due to requiring liquid nitrogen, making them largely impractical in certain scenarios, and others (consumable-free) have limited performance to modulate the lightest and most volatile analytes [35, 36]. Conversely, flow modulators utilize a redirection of gas flow to transfer eluate from the ^{1}D to ^{2}D column [37–43].

Flow modulators have become more commercially accessible in recent years and are less costly due to not requiring cryogens, making them a more feasible option for small-scale applications. However, while commercial flow modulators, like the Agilent G3486A CFT, have been successfully used with PLOT columns in both dimensions [44, 45], these are flow diversion modulators, that may require high flow rates on second dimension [29], that in turn may limit applicability of PLOT columns. Many flow modulators utilize a waste port, with resulting duty cycles of only approximately 10 to 40% [38, 46–48]. Thus, some commercial flow modulators may not be ideal for the trace analysis of low molecular weight VOCs via GC×GC, which would benefit from utilization of a full-transfer (100% duty cycle) [29, 42, 43] flow modulation technique such as dynamic pressure gradient modulation (DPGM) [41–43].

Herein, we demonstrate the capability of DPGM towards the trace analysis of lightweight VOCs in natural gas and gasoline samples. A reverse GC×GC column configuration is utilized, with a polar PLOT-BOND ^{1}D column and a non-polar PLOT-Q ^{2}D column. DPGM operates by applying an auxiliary gas pressure (P_{aux}) to a T-union, which joins the ^{1}D and ^{2}D columns, in a cyclic rhythm based on the modulation period (P_M) and pulse width (p_w) [41]. During the P_M, a suitable P_{aux} is first applied to the T-union to temporarily stop the flow of the ^{1}D eluate to the ^{2}D column. Then, P_{aux} is turned off for a time interval (p_w), allowing for the ^{1}D eluate to pass through the T-union before P_{aux} is reapplied, which rapidly reinjects the ^{1}D eluate onto the ^{2}D column; this process repeats so all ^{1}D eluate for each P_M enters the ^{2}D column in a rapid pulse [41]. Previous work has demonstrated that DPGM, can be used for a short P_M, provide full modulation (100 % duty cycle) and signal enhancement (*SE*) between 7 and 87 depending on the compound, where the *SE* is the ratio of the highest ^{2}D peak relative to the unmodulated ^{1}D peak [18–20, 41]. Application of DPGM not only provides a high *SE* that benefits the detectability of trace

compounds, but also more efficient ^{2}D separations are provided due to the production of narrow ^{2}D peaks which results in an increased in peak capacity [30]. As a stop-flow modulation technique [41], DPGM introduces a rhythmic baseline disturbance (as a function of the P_M) that is subsequently observed in the FID [47, 49]. While these baseline disturbances have been noted in the literature [50, 51], a data analysis method to correct for these baseline signal disturbances to facilitate accurate peak quantification (especially for trace analysis) has not been developed and investigated until now.

Therefore, in demonstrating the performance of DPGM for trace analysis, herein we develop a baseline correction method for GC×GC-FID using DPGM, referred to as DPGM-GC×GC-FID. Performance of this method will be evaluated by generating calibration curves for standard solutions containing three alkanes (pentane, hexane, and heptane) along with analyzing natural gas and gasoline samples. The baseline correction method will be shown to substantially extend the calibration and quantification down to the single-digit ppm range.

2.2 Experimental

All experiments were performed using DPGM-GC×GC-FID, which is based upon an Agilent 6890 GC (Agilent Technologies, Palo Alto, CA, USA) with FID as shown in Fig. 2.1. A The stock electrometer for the Agilent FID was replaced by an in-house built electrometer that allows data to be collected at 100 kHz, with the data boxcar averaged to 1 kHz. The electrometer was interfaced to a data acquisition board and the data was collected using an in-house program written in LabVIEW (National Instruments, Austin, TX, USA). Post-run data processing was performed in MATLAB R2020a (The Mathworks, Inc., Natick, MA, USA).

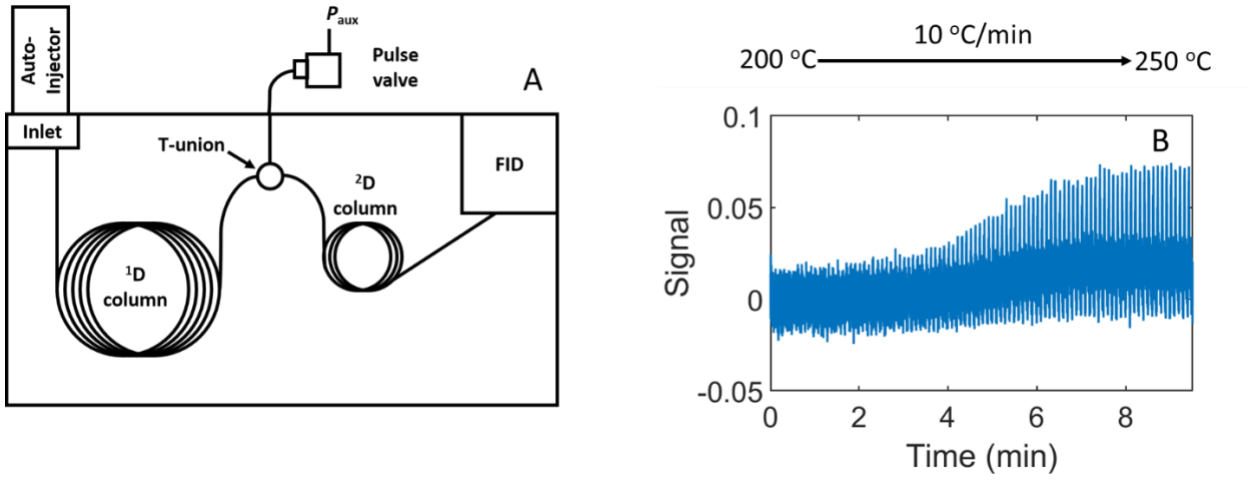


Figure 2.1. (A) Schematic of the major components of DPGM-GC×GC-FID instrument. A pulse valve provides dynamic pressure gradient modulation (DPGM) between the ^{1}D and ^{2}D columns, at the T-union by applying a suitable pressure (P_{aux}). Detection was performed with a flame ionization detector (FID). (B) Background chromatogram collected with the temperature program from 200°C to 250 °C at 10 °C /min.

A reverse GC×GC column configuration was utilized with a polar ^{1}D column (Silica BOND PLOT, 30 m × 320 µm d_c, Restek Corporation, Bellefonte, PA, USA) and a non-polar ^{2}D column (PLOT-Q, 2 m × 250 µm d_c, Restek, Bellefonte Corporation,, PA, USA). For consistency, this PLOT column configuration was used for all experiments. Constant volumetric flowrates of 1.3 ml/min and 3.9 ml/min were applied to the ^{1}D and ^{2}D columns, respectively. To implement DPGM, the ^{1}D and ^{2}D capillary GC columns were connected using a 3-way T-union (Model MT.5CXS6, Valco Instruments Company Inc., Houston, TX, USA), with the third port connected to a high-speed pulse valve (Model 009-0347-900, Parker Hannifin, Hollis, NH). The connection was achieved using a specialized fitting and a 5 cm × 125 µm i.d. stainless steel transfer line (VICI model T5C5D, Valco Instruments Company Inc., Houston, TX, USA). The pulse valve was controlled using the same previously mentioned LabVIEW program.

Initially, background chromatogram runs were obtained using $P_M = 6$ s with $p_w = 2$ s, where p_w is the time interval in which the pulse valve flow is shut off per each P_M. Ultra-high purity hydrogen (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas that was

additionally filtered using a triple gas filter (Restek Corporation, Bellefonte, PA, USA), designed to remove oxygen, moisture, and hydrocarbons. The temperature was ramped from 200 °C to 250 °C at 10 °C/min and held for 7 min, to ensure equilibration of temperature in the oven. All pressure ramp tuning for DPGM implementation was achieved as previously described [41]. The inlet pressure was ramped from 37.9 kPa (5.5 psig) to 44.1 kPa (6.4 psig) at the rate of 1.2 kPa/min (0.18 psig/min). Auxiliary pressure, P_{aux}, (i.e., independent hydrogen flow from the pulse valve) was ramped from 29.6 kPa (4.3 psig) to 34.5 kPa (5 psig) at the rate of 1 kPa/min (0.14 psig/min). The inlet and FID were set to a constant temperature of 250 °C for all experiments herein. Additionally, acetone (solvent) background runs, and background runs without sample or solvent injection, were collected at the same P_M = 6 s and p_w = 2 s one right after another (4 replicates each), to determine if the solvent influences the detected background.

A stock alkane standard solution of 20,000 ppm containing pentane, hexane, and heptane at equal concentrations (diluted in acetone) was used to develop the baseline correction method. The stock solution was diluted in acetone to create standard solutions of 4000 ppm, 2000 ppm, 1000 ppm, 500 ppm, 250 ppm, 125 ppm, 62.5 ppm, 31.3 ppm, 15.6 ppm and 7.8 ppm concentration. Sample injection was carried out using a 7683B auto-injector (Agilent Technologies, Palo Alto, CA, USA). A 0.1 µl of sample was injected without a split, for better reproducibility when quantifying the areas, and all chromatograms were acquired in quadruplicate. To reduce wraparound, a P_M = 6 s and a pulse width, p_w, of 2 s were utilized. The temperature of the GC oven was held at 200 °C and ramped after injection using a temperature program of 10 °C/min to 250 °C, where it was held for 7 min. The inlet pressure was ramped from 37.9 kPa (5.5 psig) to 44.1 kPa (6.4 psig) at the rate of 1.2 kPa/min (0.18 psig/min). Auxiliary pressure, P_{aux}, was ramped from 29.6 kPa (4.3 psig) to 34.5 kPa (5 psig) at the rate of 1 kPa/min (0.14 psig/min).

Analysis of a natural gas sample (Air Liquide, Paris, France) containing n-butane (3%), carbon dioxide (1%), ethane (9%), helium (0.5%), isobutane (3%), isopentane (1%), nitrogen (5%), n-pentane (1%) and propane (6%) in methane illustrated applicability of the baseline correction method for compositional analysis of environmental samples. A 50 µl sample was collected from the gas cylinder and manually injected, and optimal modulation conditions to reduce wraparound were utilized: P_M = 2 s, p_w = 0.55 s. The GC oven temperature was ramped from 100 ºC to 150 ºC at a rate of 10 ºC/min, then ramped from 150 ºC to 200 ºC at a rate of 20 ºC/min, and finally held at 200 °C for 5 min. The inlet pressure was first ramped from 26.9 kPa (3.9 psig) to 32.4 kPa (4.7 psig) at the rate of 1.1 kPa/min (0.16 psig/min), then ramped to 37.9 kPa (5.5psig) at a rate of 2.2 kPa/min (0.32 psig/min). Auxiliary pressure, P_{aux}, was first ramped from 20.7 kPa (3.0 psig) to 24.8 kPa (3.6 psig) at the rate of 0.8 kPa/min (0.12 psig/min), then to 29.6 kPa (4.3 psig) at a rate of 1.9 kPa/min (0.28 psig/min).

Unleaded gasoline sourced from a local gas station was analyzed to show the versatility of the separation and baseline correction methods while using a more complex sample. A 0.1 µl sample was injected in split-less mode. The analysis was performed at two different sets of modulation conditions, (1) P_M = 2 s, p_w = 0.5 s and (2) P_M = 6 s, p_w = 2 s, to illustrate both full utilization of the 2D separation space (P_M = 2 s) and minimal wraparound (P_M = 6 s). The oven temperature was ramped from 120 ºC to 230 ºC at 10 ºC/min, then from 230 ºC to 250 ºC at 20 ºC/min, and finally held constant at 250 ºC for 30 min. The inlet pressure was first ramped from 29.0 kPa (4.2 psig) to 41.4 kPa (6 psig) at the rate of 1.1 kPa/min (0.16 psig/min), then ramped to 44.1 kPa (6.4 psig) at a rate of 2.8 kPa/min (0.4 psig/min). Auxiliary pressure, P_{aux}, was first ramped from 22.1 kPa (3.2 psig) to 32.4 kPa (4.7 psig) at the rate of 1.0 kPa/min (0.14 psig/min), then to 34.5 kPa (5 psig) at a rate of 2.1 kPa/min (0.3 psig/min).

2.3 Results and Discussion

When using the DPGM-GC×GC-FID instrument (Fig. 2.1. A), the background exhibited a repetitive signal pattern that increased in intensity along the ^{1}D separation, due to the stop-flow nature of using DPGM during the temperature program at constant flow through each of the columns (Fig. 2.1. B). The most likely source of the repetitive signal pattern disturbance observed in the baseline is the flow fluctuations inherit to DPGM in conjunction with either the potential for organic compound impurities in the carrier gas or the injected acetone solvent used in sample preparation. To study the acetone solvent possibility, acetone blanks and instrument background blanks (program started using the autoinjector, but without a sample or solvent injected) were run in pairs one right after another four times in total (Fig. 2.2. A - D).

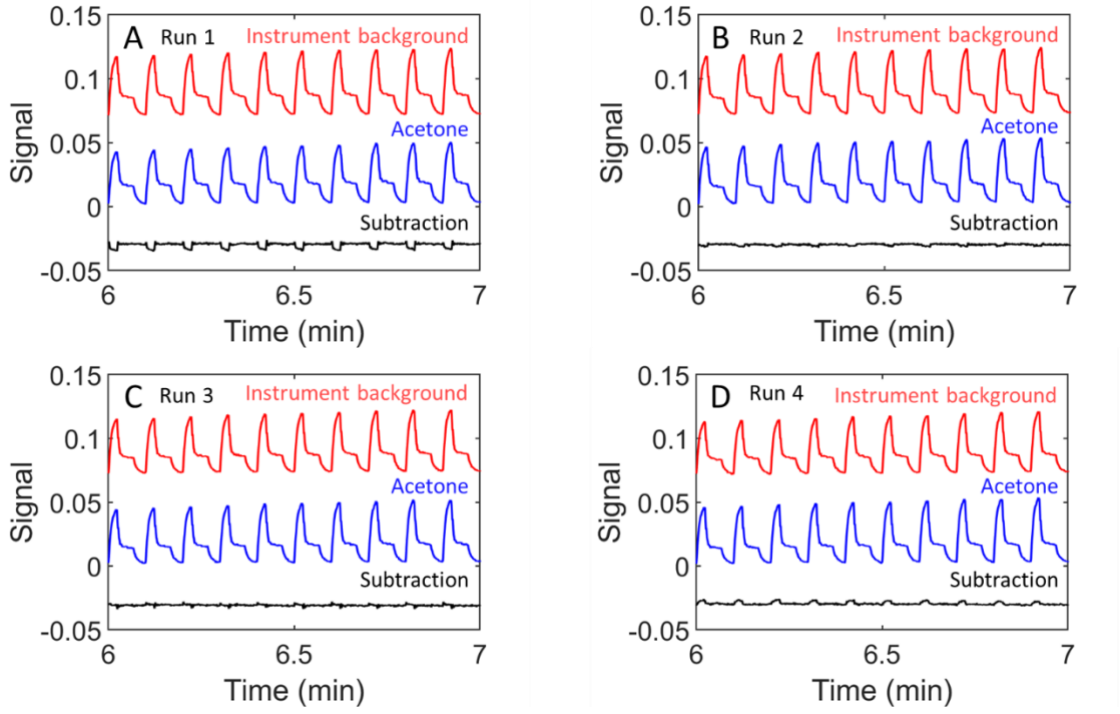


Figure 2.2. Background noise for four consecutive runs of solvent (acetone) chromatogram (blue) and instrument background (air injection) chromatogram (red). The subtraction of the instrument background from the solvent chromatogram is shown in black for runs 1-4. In blue, the acetone blank is shown, and in red, the instrument background is shown. In black, the subtraction of the instrument background from the acetone blank is presented.

The small differences indicate that the backgrounds vary slightly run-to-run, however there is also slight variation between the acetone and instrument blanks. Note that for the acetone blank runs, the large acetone peak has observed at ~ 10 min, off-scale to the right in Fig. 2.2. A-D. To examine the possibility of organic compounds in the carrier gas, we performed background runs after changing the carrier gas filter, from the initial filter to a new one, as the FID acts as a quasi-mass flow meter, which would have more ionizable atoms reaching it per unit time when the flow increases if the carrier gas contains organic compound impurities. No difference in the backgrounds was observed after the filter change. This outcome, while not totally conclusive, suggests both filters are likely removing most but not all of the organic compound impurities, and so a sufficient level of the impurities could remain to produce the observed background signal pattern. Moreover, we considered that the noise could be caused by particulate detachment from the PLOT columns during DPGM modulation, and so we repeated blank runs using a WCOT column set for both ^{1}D (CP-WAX 52CB, 30 m × 320 µm d_c × 0.2 µm d_f) and ^{2}D (RTX-5, 2 m × 250 µm d_c × 0.5 µm d_f). The same backgrounds were observed, so the PLOT columns were also dismissed as a cause for the repetitive background noise pattern. Thus, due to this signal pattern in the background, a data processing method is required to mitigate this issue in the analysis of chromatographic separations data using DPGM with the FID.

Development and illustration of the baseline correction method for the 500-ppm standard solution of pentane, hexane, and heptane (diluted in acetone) is presented in Fig. 2.3. Visual inspection of the 500-ppm standard solution separation in unfolded raw data form (Fig. 2.3A) shows that the “noise spikes” from the rhythmic baseline disturbances, which occur once each P_M, increase during the separation. This observation was reproducible injection-to-injection, if chromatographic conditions were kept constant (eg., the same temperature and flow program is

used), illustrating the need for baseline correction. A background chromatogram using the same chromatographic conditions is presented in Fig. 2.3. B.

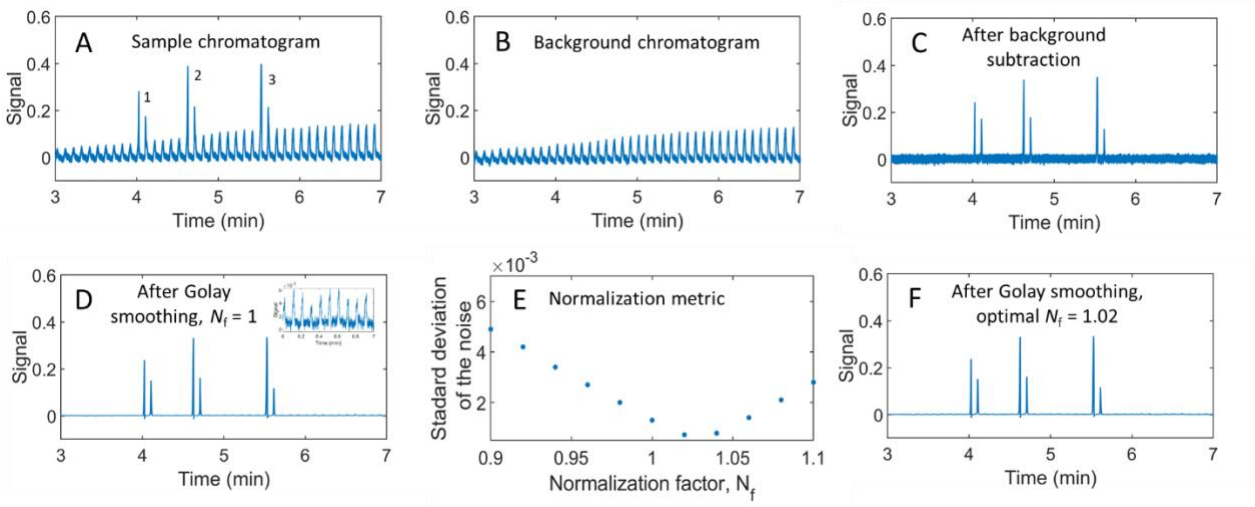


Figure 2.3. Demonstration of the baseline correction method. (A) Raw (preprocessed) modulated chromatogram of the 500-ppm alkane standard solution: (1) pentane, (2) hexane, and (3) heptane. (B) Raw background chromatogram (no sample injected). (C) Chromatogram after the background subtraction step. (D) Chromatogram after the Savitzky-Golay smoothing step. The insert shows the zoomed-in view of the noise between 6 to 7 min. (E) Standard deviation of the baseline noise in the 6 to 7 min window following background subtraction as a function of the normalization factor (N_f) applied to the background in (B) prior to subtraction. The N_f corresponding to the lowest standard deviation, is the optimized N_f. (F) Fully processed chromatogram when optimized N_f is used (N_f = 1.02).

Subtraction of the background chromatogram from the 500-ppm standard solution chromatogram in Fig. 2.3. A is shown in Fig. 2.3. C, where the baseline now appears to be flat in the sense that the noise from the rhythmic baseline disturbances has been essentially eliminated. However, high frequency noise is still present, which is why Savitzky-Golay smoothing is applied to reduce the noise even further; the resulting chromatogram after this data processing step is shown in Fig. 3D. Savitzky-Golay smoothing was performed using a 221 ms time interval for all analyses of the alkane standard solutions. The time interval selected for Savitzky-Golay smoothing was approximately 1/6th the width of a typical ^{2}D peak width (which is 1.3 s), in order to not introduce ^{2}D peak broadening due to this smoothing step. We note that following the Savitsky-Golay smoothing step, the noise spikes from the rhythmic baseline disturbances can still be seen

in the baseline, although they are much smaller in magnitude; compare the 6-7 min interval in the background chromatogram of Fig. 2.3. B to the insert in Fig. 2.3. D. This suggests that even though the noise in a given chromatogram (eg., Fig. 2.3. A) is nominally the same as that in a given background chromatogram (Fig. 2.3. B), the reality is that the noise due to the rhythmic baseline disturbances is not sufficiently close to being exactly the same. Therefore, a background normalization step was implemented to alleviate this issue (Fig. 2.3. E), analogous to using an internal standard to normalize two chromatograms relative to each other. The standard deviation of the background corrected baseline noise (after Savitzky-Golay smoothing) was calculated as a function of a normalization factor (N_f) applied using the region between 6 and 7 min, selected because no peaks eluted after ~ 6 min, except for solvent peak at 10 min. The optimal N_f is a multiplication factor that accounts for the amplitude difference between the background noise in the sample chromatogram and the background chromatogram, that for a given background chromatogram produced the lowest standard deviation of the background corrected sample chromatogram (following the smoothing of the baseline noise). For the three lowest concentrations (7.8 ppm – 31.3 ppm) the optimal N_f ranged from 0.90 to 1.02. For the 500-ppm alkane standard solution, after background normalization is performed, $N_f = 1.02$ was found to be optimal, which when applied produced a flatter baseline in Fig. 2.3. F.

While the relatively high concentration of the 500-ppm alkane standard solution is well suited for illustrating the baseline correction method in a stepwise manner, examination of a 31-ppm alkane standard solution chromatogram highlights the importance of the data processing for improving detectability. In Fig. 2.4. A, the raw data collected for the 31-ppm standard solution does not appear to contain any analyte signal from the three alkanes, and visually appears similar to the background collected (Fig. 2.4. B). After the background chromatogram is optimally

normalized and then subtracted from 31-ppm standard solution chromatogram in Fig. 2.4. A, analyte signal still cannot be distinguished from the baseline noise (Fig 2.4. C). However, after applying Savitzky-Golay smoothing, all three analytes are visible in the resulting chromatogram (Fig. 2.4. D); for example, the two ^{2}D hexane peaks, obtained from modulation from ^{1}D to ^{2}D, are readily observed demonstrating the improved detectability (Fig. 2.4. D insert).

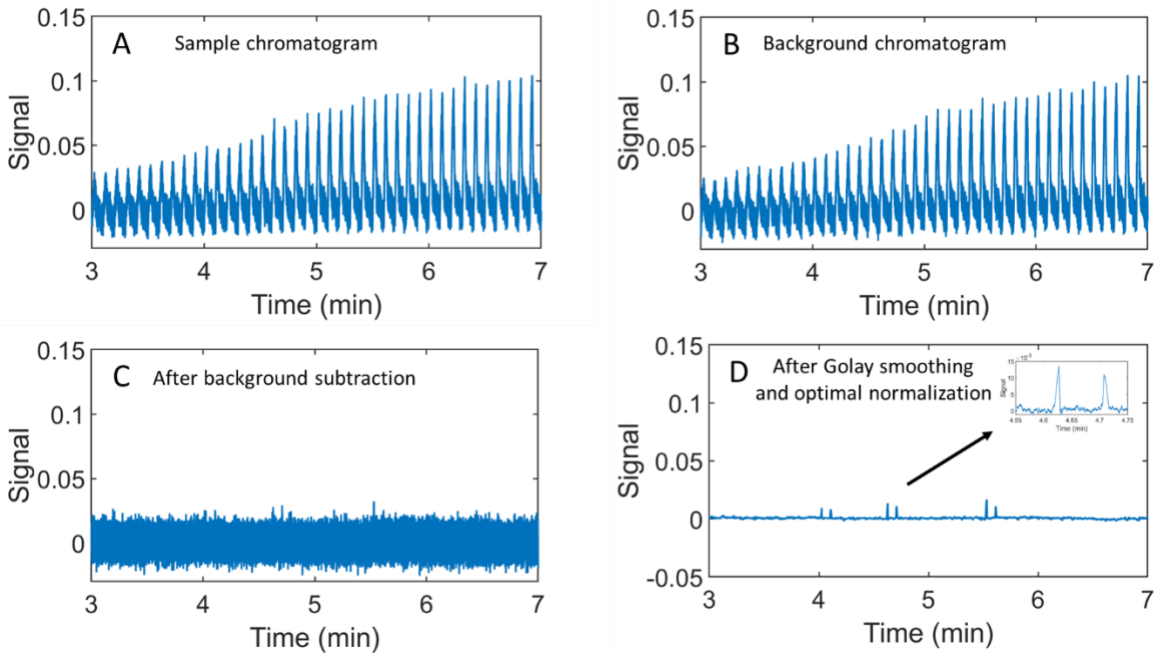


Figure 2.4. (A) Raw modulated chromatogram of the 31-ppm alkane standard solution: (1) pentane, (2) hexane, and (3) heptane. (B) Raw background chromatogram (no analyte injected). (C) Chromatogram after the background subtraction step. (D) Chromatogram after the Savitzky-Golay smoothing step when an optimized normalization factor (N_f = 1.02) is used. The insert shows the zoomed-in version of the two modulated ^{2}D hexane peaks.

To further illustrate the impact of using the normalization step during background correction (Fig. 2.3. E), the overlay of 11 N_f's $\pm$10% from N_f = 1 (by steps of 0.02) applied to 31-ppm standard solution data are shown in Fig. 2.5. Figure 2.5. A demonstrates how a baseline noise section (after background correction) changes depending upon the N_f applied, with the lowest (N_f = 0.9) producing the top blue curve and the highest (N_f = 1.1) producing the bottom purple curve. The original baseline section (N_f = 1) is highlighted in black, and the optimized background corrected baseline section (N_f = 1.02) is highlighted in green. The same color scheme was used in Fig. 2.5.

B, where the section of the chromatographic data containing the two ^{2}D hexane peaks observed (Fig. 2.4. D insert) are background corrected as a function of N_f. The importance of proper normalization is critical for peak quantification, as a non-optimal N_f can not only bias peak height and area, but also broaden and distort the peak shapes, which impacts the quality of the overall 2D chromatogram obtained. The reproducibility achieved by the baseline correction method is illustrated in Fig. 2.5. C, with overlays of the four 31-ppm standard solution replicates after the optimal N_f (ranging from 1.00 to 1.02) is applied (shifted vertically and left to right for clarity). Finally, the unfolded vector chromatograms from Fig. 2.5. C were reshaped into 2D plots of the GC×GC chromatograms, and following ^{2}D alignment, summed along ^{1}D to produce a single ^{2}D peak per analyte, with the results for the four replicates of the 31-ppm and 7.8-ppm hexane peaks presented in Fig. 2.5. D plotted along the ^{2}D time axis. Excellent reproducibility in peak height and peak width is observed.

These summed ^{2}D peaks (as in Fig. 2.5. D) were used to recreate the unmodulated peak for easier area quantification and to produce calibration curves for each analyte, with the calibration curves for hexane provided in Fig. 2.6. A, B. The full concentration range of 7.8 ppm – 4000 ppm was examined, and a linear relationship observed ($r^2 > 0.999$) in Fig. 2.6. A. Further, Fig. 2.6. B more clearly shows the calibration data for the low concentration range of 7.8 ppm – 31.3 ppm. Each data point on the calibration curves shows the average and standard deviation among four replicates, albeit the standard deviation was smaller than the dot plotted for most of the concentration levels. The corresponding calibration curves for pentane and heptane were also prepared (not shown for brevity); these analytes also exhibit linear relationships ($r^2 > 0.991$). Using the summed ^{2}D peak areas of the 7.8-ppm standard solution data, as shown in blue in Fig. 2.5. D for hexane, the limit-of-detection (*LOD*) and limit of quantification (*LOQ*) metrics were calculated

for the three alkanes in the standard solution. Hexane exhibited the best detection sensitivity, the *LOD* was 0.9 ppm, with a *LOQ* of 3.0 ppm. For pentane, the *LOD* was 2.5 ppm with a *LOQ* of 8.2 ppm. Finally, the *LOD* of heptane was determined to be 1.9 ppm, with a *LOQ* of 6.4 ppm. It is important to note that these *LOD* and *LOQ* values would have been biased and inflated if the normalization factor step shown in Fig. 2.3. E was not used. The peak areas are bigger or smaller when normalization is not used ($N_f = 1$) as in Fig. 2.6. C; the standard deviation is much higher at the lowest three concentrations compared to when the optimal N_f is used in Fig. 2.6. B.

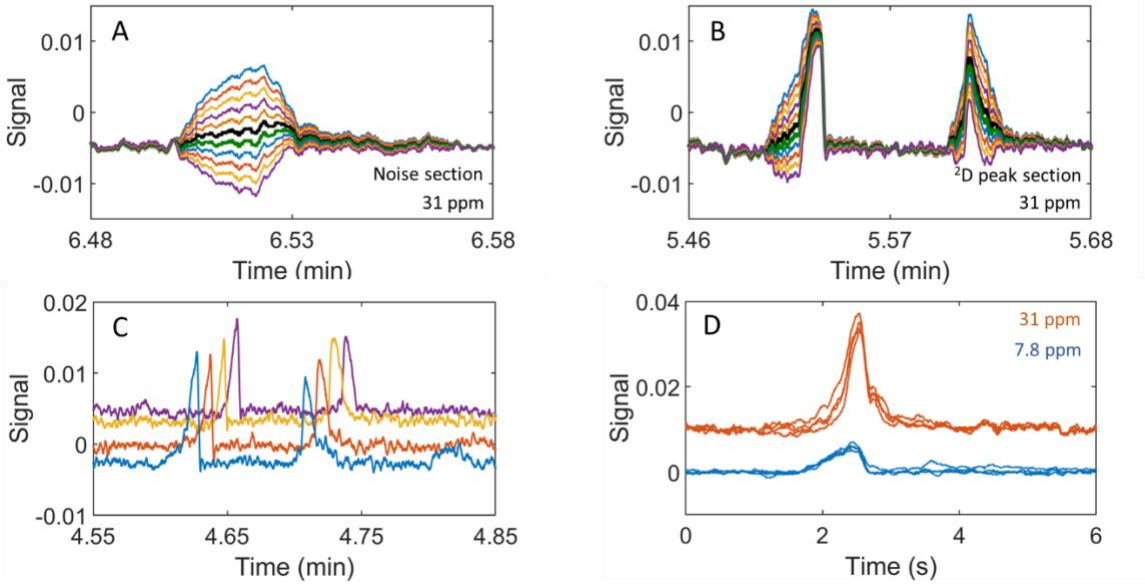


Figure 2.5. Overlay of processed vector chromatograms of 31-ppm alkane standard solution produced using different normalization factors (±10% from $N_f = 1$), where the black line represents using $N_f = 1$, and the green line represents the optimal $N_f = 1.02$. (A) Baseline noise section (no analytes eluting). (B) Section of the data where hexane ^{2}D peaks elute. (C) Unfolded data of 31-ppm hexane standard solution for 4 replicates, after data processing using an optimized N_f ranging from 1.00 to 1.02. (D) Summed ^{2}D hexane peaks obtained by summing along ^{1}D for 4 replicates at 31 ppm and 7.8 ppm.

The benefit of using normalization is further seen by comparing the percent deviation (% deviation) in the peak area in Fig. 2.6. D, between when $N_f = 1$ is used (Fig. 2.6. C) (in other words, when the normalization step is omitted), and when N_f = optimal is used (Fig. 2.6. B). Here the % deviation at a given standard concentration is defined as the peak area obtained from an optimally normalized peak area (optimal N_f) subtracted from the peak area at $N_f = 1$ (Fig. 2.6.

C) divided by the peak area obtained from optimally normalized peak area, then multiplied by 100 percent, averaged for the replicates. The % deviation is smaller (essentially 0%) at higher concentrations but is substantially less at 15.6 ppm and 7.8 ppm when background normalization is applied. Indeed, when background normalization is omitted ($N_f = 1$), the areas obtained varied by as much as 120%. Based on the results from Fig. 2.6. we conclude that the background normalization correction improved analyte detectability, which in turn benefits the *LOD* and *LOQ*. However, the intent of this work is not to compare the *LOD* and *LOQ* obtained to other methods, but rather to improve the performance of DPGM-GC×GC-FID.

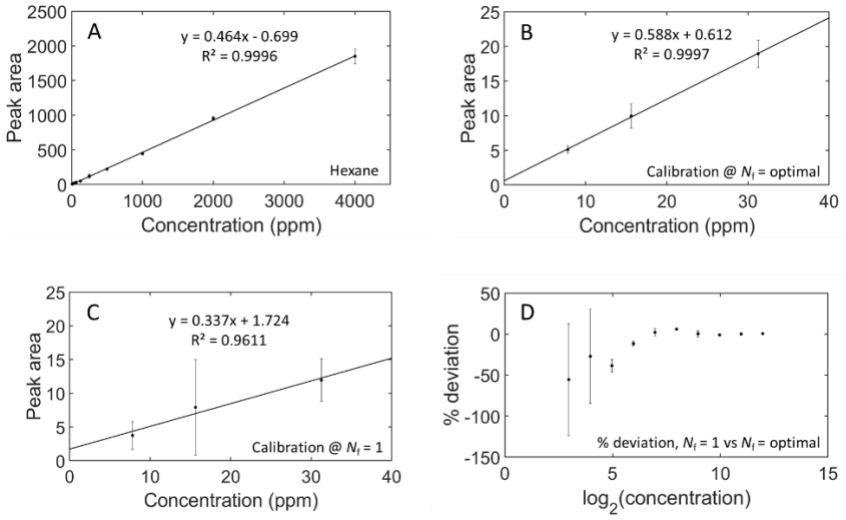


Figure 2.6. (A) Peak area versus concentration calibration plot for hexane (7.8 ppm – 4000 ppm). Peak areas measured from summed ^2D peaks as in Fig. 5D. Calibration plots for lowest three concentrations (7.8 ppm, 15.6 ppm, 31.3 ppm) with optimal N_f (B) and with $N_f = 1$ (C). (D) Percent deviation (% deviation) between peak areas obtained when N_f = optimal and when $N_f = 1$ for the full concentration range 7.8 ppm – 4000 ppm on a $\log_2$ concentration scale.

A separation of a natural gas sample in Fig. 2.7. illustrates the efficacy of the baseline correction method for a real sample. Since natural gas is a combination of mostly methane, with only small amounts of other volatile hydrocarbons, accurate trace analysis benefits greatly from the method. The GC×GC-FID data in unfolded vector form of the natural gas sample after application of the method is shown in Fig. 2.7. A, where all the hydrocarbons have baseline

resolution including the butane and pentane isomers. Application of the method allows for detection of both isopentane and n-pentane, that each compose only 1% of the natural gas sample. The folded GC×GC-FID chromatogram (Fig. 2.7. B) provides adequate separation of these C1-C5 alkanes along ^{1}D, while on ^{2}D they are minimally retained. However, and more importantly, application of DPGM-GC×GC-FID provides 100% duty cycle modulation, resulting in appreciable detection sensitivity enhancement for the analysis of the low concentration analytes in the natural gas sample (Fig. 2.7. C, D). The signal enhancement (*SE*) is calculated by dividing the height of the tallest ^{2}D peak in the modulated chromatogram for the analyte by the height of the unmodulated ^{1}D peak [18–20, 41], providing a *SE* of 30 for isopentane (Fig. 2.7. C, D), as the ^{1}D peak width-at-base is about 10 s, modulated about 5 times (with P_M = 2 s) into ^{2}D peaks with a width-at-base of 70 ms. This is particularly important for isopentane and n-pentane, as these analytes present in the natural gas sample have been known to be difficult to quantify due to their low concentration and low *S/N* [10]. Indeed, the tallest modulated isopentane ^{2}D peaks (Fig. 2.7. C insert) has an 11-fold higher *S/N* than the boxcar averaged unmodulated ^{1}D peak (Fig. 2.7. D white trace) for isopentane (resampled to have the same number of data points for an objective comparison). Additionally, we confirmed that the total peak area of the modulated ^{2}D peaks was equal to the area of the ^{1}D peak (within 3%), confirming a 100% duty cycle.

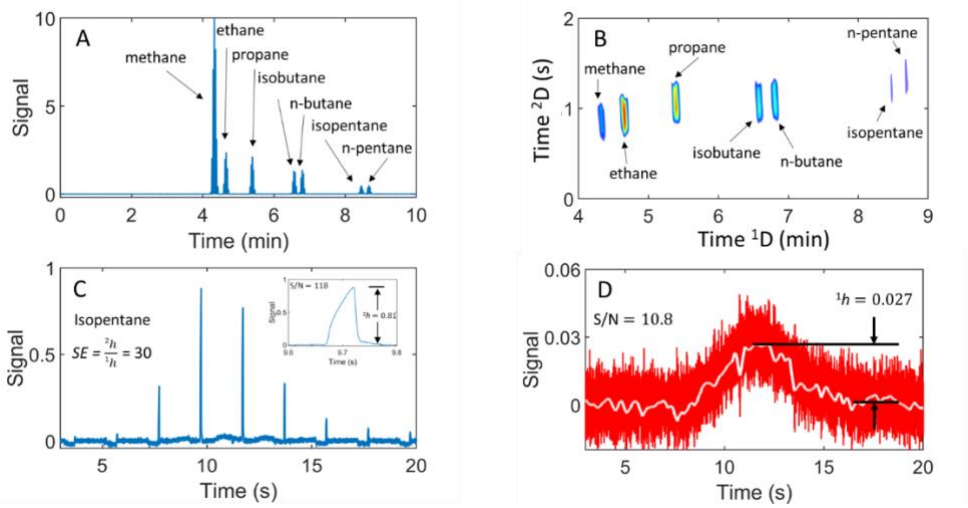


Figure 2.7. Analysis of a natural gas sample. (A) Unfolded vector form chromatogram of a natural gas sample separation, following the baseline correction method, showing full separation of light hydrocarbons in the sample. (B) Processed GC×GC chromatogram that has made from (A) by folding the data at the PM. Methane peak was purposely plotted on a 10-fold less sensitive scale, to more accurately reveal less concentrated peaks in the sample. (C) The modulated isopentane 2D peaks with the signal enhancement (SE) calculation illustrated. The insert shows the zoom-in of the tallest 2D peak. (D) Unmodulated isopentane 1D peak (red) with a white trace showing the unmodulated peak that has been boxcar averaged to contain the same number of data points as the tallest modulated 2D peak in (C). In both (C) and (D) the retention times shown are relative; the pentane peak in (C) elutes about 2 min later than in (D).

The DPGM-GC×GC-FID chromatograms of the unleaded gasoline sample analyzed with P_M = 2 s and P_M = 6 s (after applying the method) and P_M = 6 s (without applying the method) are provided in Fig. 2.8. A-C, respectively, to provide an example of a complex sample analyzed with the baseline correction method in which several of the analyte peaks are not baseline resolved. The two modulation periods produced similar 2D peak capacities (Fig. 2.8. A, B): $n_{c,2D}$ = 231 for P_M = 2 s, and $n_{c,2D}$ = 250 for P_M = 6 s. Although the ^{1}D peak widths on average were nearly identical, the average ^{2}D peak widths were found to be proportionate to the modulation period: ^1W_b = 26 s and ^2W_b = 0.8 s for P_M = 2 s, and ^1W_b = 24 s and ^2W_b = 2.4 s for P_M = 6 s. Although it is not obvious in the separations, a P_M = 6 s avoided most of the wraparound of the ^{2}D separations, while the P_M = 2 s allowed the ^{2}D separations to wrap around a couple times. The longer modulation period of P_M = 6 s provided a shorter ^{1}D run time, which is a consequence of the stop-flow nature of applying DPGM. More modulations per ^{1}D run time slows that separation

to a greater extent. Pentane, hexane, and heptane were identified by retention time comparison to the injection of a standard solution and are labeled accordingly in Fig. 2.8. A, B. Without application of the data analysis method to remove the background noise, the smaller, less concentrated peaks that appear in the chromatogram after 25 min would not be visible (Fig. 2.8. C), which highlights the importance of the baseline correction method for complex samples. It also means that the data processing has the most impact on the compounds that elute at high temperature and are unretained on the second dimension (Fig 2.8. B, C).

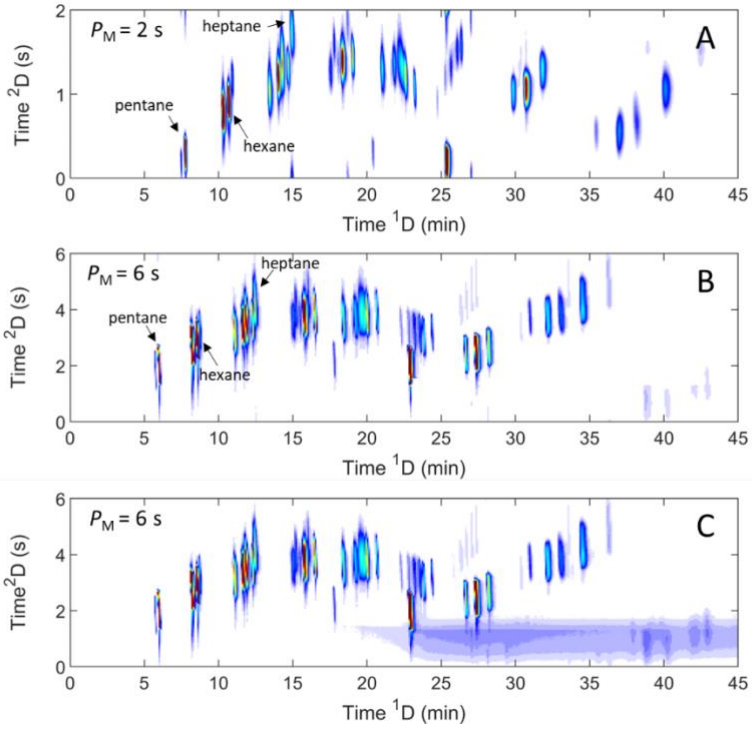


Figure 2.8. Analysis of an unleaded gasoline sample. GC×GC chromatograms are produced for the separation at (A) P_M = 2 s, p_w = 0.5 s and (B) P_M = 6 s, p_w = 2 s, with the baseline correction method applied in both cases. (C) GC×GC chromatogram is produced for the separation at P_M = 6 s, p_w = 2 s without applying the baseline correction method. After 25 min, darker feature is due to the baseline noise.

2.4 Conclusion

A baseline correction method for DPGM-GC×GC-FID analysis of trace level VOCs has been presented. Method development was demonstrated via analysis of alkane standard solutions (containing pentane, hexane, and heptane) and obtained linear calibration curves (r^2 > 0.991) for

the three compounds. *LOD* and *LOQ* metrics were calculated for pentane (*LOD* = 2.5 ppm, *LOQ* = 8.2 ppm), hexane (*LOD* = 0.9 ppm, *LOQ* = 3.0 ppm), and heptane (*LOD* = 1.9 ppm, *LOQ* = 6.4 ppm). Analysis of a natural gas sample examined the utility of the overall instrumental and data analysis method for analytes below C6. Key analytes in the natural gas sample were baseline separated, whereby the *S/N* for isopentane increased 11-fold with an *SE* of 30. Improved detection sensitivity is also demonstrated for a gasoline sample, where a much cleaner GC×GC-FID chromatogram is obtained after the baseline correction method is applied. Future work on optimizing instrumental design to gain the ability to obtain background and sample chromatograms at the same time could in principle be performed. However, our experience was that by collecting a representative background chromatogram once per day, that the other sample chromatograms could be adequately corrected by the method presented. Thus, the method broadens the scope and utility of DPGM-GC×GC-FID to trace analysis applications.

2.5 References

[1] V.I. Safarova, S.V. Sapelnikova, E.V. Djazhenko, G.I. Teplova, G.F. Shajdulina, F.K. Kudasheva, Gas chromatography–mass spectrometry with headspace for the analysis of volatile organic compounds in waste water, J. Chromatogr. B 800 (2004) 325–330. https://doi.org/10.1016/j.jchromb.2003.10.070.

[2] C.V. Antoniou, E.E. Koukouraki, E. Diamadopoulos, Determination of chlorinated volatile organic compounds in water and municipal wastewater using headspace–solid phase microextraction–gas chromatography, J. Chromatogr. A 1132 (2006) 310–314. https://doi.org/10.1016/j.chroma.2006.08.082.

[3] C. Brasseur, J. Dekeirsschieter, E.M.J. Schotsmans, S. de Koning, A.S. Wilson, E. Haubruge, J.-F. Focant, Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry for the forensic study of cadaveric volatile organic compounds released in soil by buried decaying pig carcasses, J. Chromatogr. A 1255 (2012) 163–170. https://doi.org/10.1016/j.chroma.2012.03.048.

[4] A. Serrano, M. Gallego, Sorption study of 25 volatile organic compounds in several Mediterranean soils using headspace–gas chromatography–mass spectrometry, J. Chromatogr. A 1118 (2006) 261–270. https://doi.org/10.1016/j.chroma.2006.03.095.

[5] S.A. Baylis, K. Hall, E.J. Jumeau, The analysis of the C1–C5 components of natural gas samples using gas chromatography-combustion-isotope ratio mass spectrometry, Org. Geochem. 21 (1994) 777–785. https://doi.org/10.1016/0146-6380(94)90019-1.

[6] L.S. Moreira, L.A. d'Avila, D.A. Azevedo, Automotive Gasoline Quality Analysis by Gas Chromatography: Study of Adulteration, Chromatographia 58 (2003) 501–505. https://doi.org/10.1365/s10337-003-0065-z.

[7] P. Doble, M. Sandercock, E. Du Pasquier, P. Petocz, C. Roux, M. Dawson, Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks, Forensic Sci. Int. 132 (2003) 26–39. https://doi.org/10.1016/S0379-0738(03)00002-1.

[8] P.E. Sudol, K.M. Pierce, S.E. Prebihalo, K.J. Skogerboe, B.W. Wright, R.E. Synovec, Development of gas chromatographic pattern recognition and classification tools for compliance and forensic analyses of fuels: A review, Anal. Chim. Acta 1132 (2020) 157–186. https://doi.org/10.1016/j.aca.2020.07.027.

[9] L. Bai, J. Smuts, P. Walsh, H. Fan, Z. Hildenbrand, D. Wong, D. Wetz, K.A. Schug, Permanent gas analysis using gas chromatography with vacuum ultraviolet detection, J. Chromatogr. A 1388 (2015) 244–250. https://doi.org/10.1016/j.chroma.2015.02.007.

[10] A.S. Brown, M.J.T. Milton, C.J. Cowper, G.D. Squire, W. Bremser, R.W. Branch, Analysis of natural gas by gas chromatography: Reduction of correlated uncertainties by normalisation, J. Chromatogr. A 1040 (2004) 215–225. https://doi.org/10.1016/j.chroma.2004.04.007.

[11] M.A. Buldakov, V.A. Korolkov, I.I. Matrosov, D.V. Petrov, A.A. Tikhomirov, B.V. Korolev, Analyzing natural gas by spontaneous Raman scattering spectroscopy, J. Opt. Technol. 80 (2013) 426–430. https://doi.org/10.1364/JOT.80.000426.

[12] A. Petculescu, R.M. Lueptow, Quantitative Acoustic Relaxational Spectroscopy for real-time monitoring of natural gas: A perspective on its potential, Sens. Actuators B Chem. 169 (2012) 121–127. https://doi.org/10.1016/j.snb.2012.03.086.

[13] M.M. Rahman, A.M.A. El-Aty, J.-H. Choi, H.-C. Shin, S.C. Shin, J.-H. Shim, Basic Overview on Gas Chromatography Columns, in: Anal. Sep. Sci., American Cancer Society, 2015: pp. 823–834. https://doi.org/10.1002/9783527678129.assep024.

[14] D. Bennett, Analysis of gas mixtures by gas chromatography, J. Chromatogr. A 26 (1967) 482–484. https://doi.org/10.1016/S0021-9673(01)98907-1.

[15] J.S. Stufkens, H.J. Bogaard, Rapid method for the determination of the composition of natural gas by gas chromatography, Anal. Chem. 47 (1975) 383–386. https://doi.org/10.1021/ac60353a060.

[16] R. Gras, J. Luong, V. Carter, L. Sieben, H. Cortes, Practical method for the measurement of Alkyl mercaptans in natural gas by multi-dimensional gas chromatography, capillary flow technology, and flame ionization detection, J. Chromatogr. A 1216 (2009) 2776–2782. https://doi.org/10.1016/j.chroma.2008.09.029.

[17] J.L. Burger, T.M. Lovestead, T.J. Bruno, Composition of the C6+ Fraction of Natural Gas by Multiple Porous Layer Open Tubular Capillaries Maintained at Low Temperatures, Energy Fuels 30 (2016) 2119–2126. https://doi.org/10.1021/acs.energyfuels.6b00043.

[18] J. Krupčík, P. Májek, R. Gorovenko, I. Špánik, P. Sandra, D.W. Armstrong, On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1286 (2013) 235–240. https://doi.org/10.1016/j.chroma.2013.02.068.

[19] T.J. Trinklein, S. Schöneich, P.E. Sudol, C.G. Warren, D.V. Gough, R.E. Synovec, Total-transfer comprehensive three-dimensional gas chromatography with time-of-flight mass

spectrometry, J. Chromatogr. A 1634 (2020) 461654. https://doi.org/10.1016/j.chroma.2020.461654.

[20] A.L. Lee, K.D. Bartle, A.C. Lewis, A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography, Anal. Chem. 73 (2001) 1330–1335. https://doi.org/10.1021/ac001120s.

[21] Z. Liu, J.B. Phillips, Comprehensive Two-Dimensional Gas Chromatography using an On-Column Thermal Modulator Interface, J. Chromatogr. Sci. 29 (1991) 227–231. https://doi.org/10.1093/chromsci/29.6.227.

[22] B. Gruber, B.A. Weggler, R. Jaramillo, K.A. Murrell, P.K. Piotrowski, F.L. Dorman, Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends, TrAC Trends Anal. Chem. 105 (2018) 292–301. https://doi.org/10.1016/j.trac.2018.05.017.

[23] H.P. Tan, T.S. Wan, C.L.S. Min, M. Osborne, K.H. Ng, Quantitative analysis of fragrance in selectable one dimensional or two dimensional gas chromatography-mass spectrometry with simultaneous detection of multiple detectors in single injection, J. Chromatogr. A 1333 (2014) 106–115. https://doi.org/10.1016/j.chroma.2014.01.073.

[24] L. Mondello, A. Casilli, P.Q. Tranchida, G. Dugo, P. Dugo, Comprehensive two-dimensional gas chromatography in combination with rapid scanning quadrupole mass spectrometry in perfume analysis, J. Chromatogr. A 1067 (2005) 235–243. https://doi.org/10.1016/j.chroma.2004.09.040.

[25] E. Skoczyńska, P. Korytár, J. de Boer, Maximizing Chromatographic Information from Environmental Extracts by GCxGC-ToF-MS, Environ. Sci. Technol. 42 (2008) 6611–6618. https://doi.org/10.1021/es703229t.

[26] L.R. Bordajandi, J.J. Ramos, J. Sanz, M.J. González, L. Ramos, Comprehensive two-dimensional gas chromatography in the screening of persistent organohalogenated pollutants in environmental samples, J. Chromatogr. A 1186 (2008) 312–324. https://doi.org/10.1016/j.chroma.2007.12.013.

[27] Y. Nolvachai, C. Kulsing, P.J. Marriott, Multidimensional gas chromatography in food analysis, TrAC Trends Anal. Chem. 96 (2017) 124–137. https://doi.org/10.1016/j.trac.2017.05.001.

[28] M. Edwards, A. Mostafa, T. Górecki, Modulation in comprehensive two-dimensional gas chromatography: 20 years of innovation, Anal. Bioanal. Chem. 401 (2011) 2335–2349. https://doi.org/10.1007/s00216-011-5100-6.

[29] H.D. Bahaghighat, C.E. Freye, R.E. Synovec, Recent advances in modulator technology for comprehensive two dimensional gas chromatography, TrAC Trends Anal. Chem. 113 (2019) 379–391. https://doi.org/10.1016/j.trac.2018.04.016.

[30] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A 1383 (2015) 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.

[31] J. Luong, X. Guan, S. Xu, R. Gras, R.A. Shellie, Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 88 (2016) 8428–8432. https://doi.org/10.1021/acs.analchem.6b02525.

[32] T. Hyötyläinen, M. Kallio, K. Hartonen, M. Jussila, S. Palonen, M.-L. Riekkola, Modulator Design for Comprehensive Two-Dimensional Gas Chromatography: Quantitative

Analysis of Polyaromatic Hydrocarbons and Polychlorinated Biphenyls, Anal. Chem. 74 (2002) 4441–4446. https://doi.org/10.1021/ac0201528.

[33] A. Mostafa, T. Górecki, Development and Design of a Single-Stage Cryogenic Modulator for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 88 (2016) 5414–5423. https://doi.org/10.1021/acs.analchem.6b00767.

[34] V. Mucédola, L.C.S. Vieira, D. Pierone, A.L. Gobbi, R.J. Poppi, L.W. Hantao, Thermal desorption modulation for comprehensive two-dimensional gas chromatography using a simple and inexpensive segmented-loop fluidic interface, Talanta 164 (2017) 470–476. https://doi.org/10.1016/j.talanta.2016.12.005.

[35] W.-C. Liao, C.-F. Ou-Yang, C.-H. Wang, C.-C. Chang, J.-L. Wang, Two-dimensional gas chromatographic analysis of ambient light hydrocarbons, J. Chromatogr. A 1294 (2013) 122–129. https://doi.org/10.1016/j.chroma.2013.04.008.

[36] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

[37] J.V. Seeley, F. Kramp, C.J. Hicks, Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation, Anal. Chem. 72 (2000) 4346–4352. https://doi.org/10.1021/ac000249z.

[38] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1424 (2015) 127–133. https://doi.org/10.1016/j.chroma.2015.10.098.

[39] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.

[40] C.E. Freye, H.D. Bahaghighat, R.E. Synovec, Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period, Talanta 177 (2018) 142–149. https://doi.org/10.1016/j.talanta.2017.08.095.

[41] T.J. Trinklein, D.V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1609 (2020) 460488. https://doi.org/10.1016/j.chroma.2019.460488.

[42] S. Schöneich, D.V. Gough, T.J. Trinklein, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection, J. Chromatogr. A 1620 (2020) 460982. https://doi.org/10.1016/j.chroma.2020.460982.

[43] S. Schöneich, T.J. Trinklein, C.G. Warren, R.E. Synovec, A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations, Anal. Chim. Acta 1134 (2020) 115–124. https://doi.org/10.1016/j.aca.2020.08.023.

[44] A.Y. Sholokhova, Y.V. Patrushev, V.N. Sidelnikov, A.K. Buryak, Analysis of light components in pyrolysis products by comprehensive two-dimensional gas chromatography with PLOT columns, Talanta 209 (2020) 120448. https://doi.org/10.1016/j.talanta.2019.120448.

[45] Y.V. Patrushev, V.N. Sidelnikov, Selection of the porous layer open tubular columns for separation of light components in comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1579 (2018) 83–88. https://doi.org/10.1016/j.chroma.2018.10.015.

[46] C.A. Bruckner, B.J. Prazen, R.E. Synovec, Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis, Anal. Chem. 70 (1998) 2796–2804. https://doi.org/10.1021/ac980164m.

[47] P.Q. Tranchida, S. Salivo, F.A. Franchina, L. Mondello, Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography Combined with a High-Resolution Time-of-Flight Mass Spectrometer: A Proof-of-Principle Study, Anal. Chem. 87 (2015) 2925–2930. https://doi.org/10.1021/ac5044175.

[48] F. Stilo, E. Gabetti, C. Bicchi, A. Carretta, D. Peroni, S.E. Reichenbach, C. Cordero, J. McCurry, A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods, J. Chromatogr. A 1627 (2020) 461396. https://doi.org/10.1016/j.chroma.2020.461396.

[49] N. Oldridge, O. Panic, T. Górecki, Stop-flow comprehensive two-dimensional gas chromatography with pneumatic switching, J. Sep. Sci. 31 (2008) 3375–3384. https://doi.org/10.1002/jssc.200800265.

[50] R.E. Mohler, B.J. Prazen, R.E. Synovec, Total-transfer, valve-based comprehensive two-dimensional gas chromatography, Anal. Chim. Acta 555 (2006) 68–74. https://doi.org/10.1016/j.aca.2005.08.072.

[51] X. Guan, J. Luong, Z. Yu, H. Jiang, Quasi-Stop-Flow Modulation Strategy for Comprehensive Two-Dimensional Gas Chromatography, Anal. Chem. 92 (2020) 6251–6256. https://doi.org/10.1021/acs.analchem.0c00814.

Chapter 3. Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension

This chapter is reproduced from Lina Mikaliunaite, Timothy J. Trinklein, Grant S. Ochoa, Paige E. Sudol, David S. Bell, Robert E. Synovec, "Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension" *J. Chromatogr. O* 3 (2023) 100076.

3.1 Introduction

Comprehensive two-dimensional (2D) gas chromatography (GC×GC) [1] has gained significant popularity as an analysis tool for complex samples in metabolomics [2–4], cosmetics [5], petroleum [6,7], and others [8–11]. GC×GC has been shown to perform better than one-dimensional GC, including a 10-fold increase in peak capacity [12] and more chemical selectivity. The hyphenation of the peak capacity afforded by GC×GC with mass spectrometry detection (GC×GC-MS) creates a powerful coupling, well suited for targeted and untargeted analyses in a variety of fields [13,14]. The objective of our current study is to design and evaluate a low-cost GC×GC instrument with quadrupole MS detection (GC×GC-qMS) for the analysis of samples containing an extended volatility range from highly volatile analytes to analytes of more moderate volatility.

Analysis of volatile hydrocarbons (<C5) in complex samples by GC×GC-MS is challenging. First, wall coated open tubular columns (WCOT) are not retentive for highly volatile hydrocarbons, such as permanent gases and very small hydrocarbons [15]. Second, thermal modulators have limited ability to trap highly volatile compounds and do not trap permanent gases [16]. To address the first challenge, we use a porous layer open tubular (PLOT) column in the first dimension (^{1}D). We demonstrate the effectiveness of the ^{1}D PLOT column in combination with a polar WCOT column on ^{2}D to separate analytes across a broad volatility range. Generally, PLOT

columns have much higher retention than WCOT columns, which provides a challenge to adequately use the 2D separation space with GC×GC. One approach to address this issue would be to use a substantially elevated temperature program for the ^{1}D PLOT column relative to the ^{2}D WCOT column. However, this approach is not practical given the temperature limits of PLOT columns. Instead, we present an instrumental design that allows for the use of a single oven by increasing the ^{2}D WCOT column retention by using a relatively long column with a custom thicker WCOT film. In concert, to mitigate the volatility limit of thermal modulators, a high temperature diaphragm valve (HT-DV) modulator is used, which modulates analytes independent of volatility, including permanent gases. A schematic of the instrument is shown in Fig. 3.1.

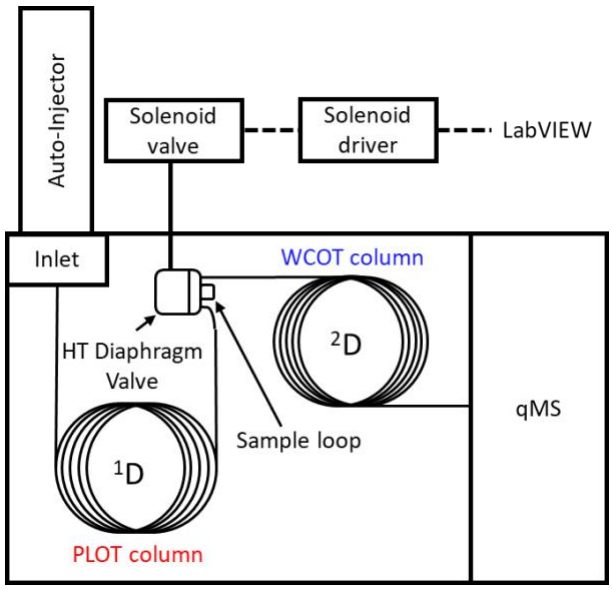


Figure 3.1. (A) Schematic of the major components of the GC×GC-qMS instrument. A high temperature diaphragm valve (HT-DV) with a 10 µL sample loop is used as a modulator between the first-dimension (1D) and second-dimension (2D) columns. Detection was performed with a quadrupole mass spectrometer (qMS).

Improvements in both PLOT column and diaphragm valve technology have made this combination more feasible than in previous years. However, reports of GC×GC employing PLOT columns remain relatively uncommon [16,17]. Early PLOT columns were made with submicron particles (e.g., aluminum oxide) adhered to the inner walls of a fused silica capillary via static

coating [18]. Although excellent separations for light analytes were achieved, the detachment of particles from the column was an issue. As a result, analysts were hesitant to use PLOT columns with MS detection for fear of damaging the instrumentation. Due to recent improvements in column technology, adsorbent layers can now be immobilized very effectively to capillary tubing making the use of PLOT columns with MS detection more appealing [19].

Likewise, diaphragm valve modulation technology has greatly improved. Modern valve designs have a high duty cycle (>80%) and no longer have a low temperature limit. The temperature restriction of early diaphragm valves (≤175 °C) was due to the decomposition of the O-rings. By implementing O-rings made with a perfluoroelastomer (Kalrez), long-term operation of diaphragm valves at up to 325 °C has been demonstrated with GC×GC using either both flame ionization detection (FID) [20] or MS [21], i.e., as a high temperature diaphragm valve (HT-DV) modulator. A unique feature and major benefit of using a diaphragm valve over other flow modulators that have been previously used with PLOT columns, like Dean's switch [22], is that the flows on ^{1}D and ^{2}D are completely decoupled with the HT-DV modulator. Decoupling of the ^{1}D and ^{2}D flows is important when using qMS detection, as it has maximum flow limitation. Herein, HT-DV modulation allowed use of a low flow of 2.0 mL/min on the ^{2}D column, which is within optimal range for qMS detection. When changing columns, which necessitates pumping down the mass spectrometer, the large changes in flow can cause particles on certain types of PLOT columns to dislodge. However, since the PLOT column can be installed on ^{1}D, and is decoupled from the ^{2}D column when the system is pumped down, any possible particles dislodged from the PLOT column would be discharged into the waste line, rather than the MS. Other flow modulator designs, such as those based on Dean's switching [16,22], the reverse fill-flush

modulator [23], or others [17,24,25], are also well suited for hyphenation with PLOT columns, especially when state-of-the-art PLOT columns are used.

Due to the narrow peak widths generated on ^{2}D with GC×GC (~100 to 300 ms), a fast scan rate is required for adequate sampling by the detector. Therefore, time-of-flight mass spectrometry (TOFMS) is recognized as a gold standard mass analyzer for GC×GC, with scan rates up to 500 Hz. In contrast, qMS generally has a much lower scan rate, typically in the range of about 10 to 30 Hz. Unlike TOFMS, the scan rate for qMS depends on the mass range selected by the analyst. However, qMS instruments are significantly less expensive, more compact, and highly robust. Because of their attractive qualities, significant research has been directed to the development and evaluation of GC×GC-qMS [26–28].

However, another issue remains when developing GC×GC-qMS methods which has received surprisingly little attention in the literature. The time required to obtain a mass spectrum, here referred to as the cycle time, is a function of the mass range scanned, and is the inverse of the scan rate. If the cycle time does not divide evenly into the modulation period, P_M, visual distortions will appear in the GC×GC chromatogram. When the cycle time does not divide evenly into the P_M, the magnitude of the remainder will dictate the extent of visual distortion. Interestingly, this effect was observed in the first GC×GC-MS report, which used a qMS detector [29]. Despite early recognition of this problem, no solutions were reported in the literature, albeit it is possible that computational methods to correct this issue are present in commercial GC×GC software packages. Unfortunately, the issue remains for analysts using in-house software to process their data. Very recently, MATLAB code to address this problem via resampling was released by Hantao and coworkers [30]. This method resamples the data to a new "effective" cycle time that divides evenly into the P_M, thus rendering any scan rate suitable for qMS detection with GC×GC. Here, we expand

on the work of Hantao and provide an in-depth explanation of the issue, in the context of the HT-DV modulated GC×GC-qMS instrument. We describe and provide both experimental and computational methods to address the issue of visual distortion in the GC×GC chromatograms with suitable solutions.

The HT-DV modulated GC×GC-qMS instrument is initially evaluated via the separation of gasoline, demonstrating excellent separation of highly volatile compounds. Next, the resampling method to correct the distortions when plotting GC×GC-qMS data is presented using the gasoline separation. Then, we examine the separation of a complex standard mixture that has a wide boiling point range of analytes, thus containing a relatively wide range of analyte volatility from highly volatile to moderate volatility. Finally, a JP8 fuel was analyzed to assess quantitative reproducibility and limit-of-detection (*LOD*) of this HT-DV modulated GC×GC-qMS instrument. Finally, implications for broad applicability of the instrument are discussed.

3.2 Theory

In quadrupole mass spectrometry, the time to acquire a mass spectrum is a sum of the scan time and the interscan delay, defined here as the cycle time:

$$\text{cycle time (s)} = \text{scan time (s)} + \text{interscan delay (s)} \tag{1}$$

The scan time is given by

$$\text{scan time (s)} = \frac{\text{mass range (amu)}}{\text{scan speed } \left(\frac{\text{amu}}{\text{s}}\right)} \tag{2}$$

during the scan time, both the DC and RF voltages of the quadrupole are ramped linearly with time, where the scan rate is a parameter defined by the instrument. Then, during the interscan delay, the DC and RF voltages are dropped and held constant. This process is shown in Fig. 3.2. A

with the variables in Eq. (1) indicated. The inverse of the cycle time determines the scan rate, which is the number of spectra acquired in Hz,

$$\text{scan rate (Hz)} = \frac{1}{\text{cycle time (s)}} \tag{3}$$

Maximizing the scan rate is critical to adequately sample the narrow ^{2}D peaks generated in GC×GC. In order to do so, the scan time and interscan delay should be minimized. The scan time is a function of the number of amu scanned, *i.e.*, the *m/z* range scanned. Therefore, a simple way to minimize the scan time and increase the scan rate is to minimize the mass range scanned. Measured on the qMS used herein, the cycle time as a function of mass range is given in Fig. 3.2. B.

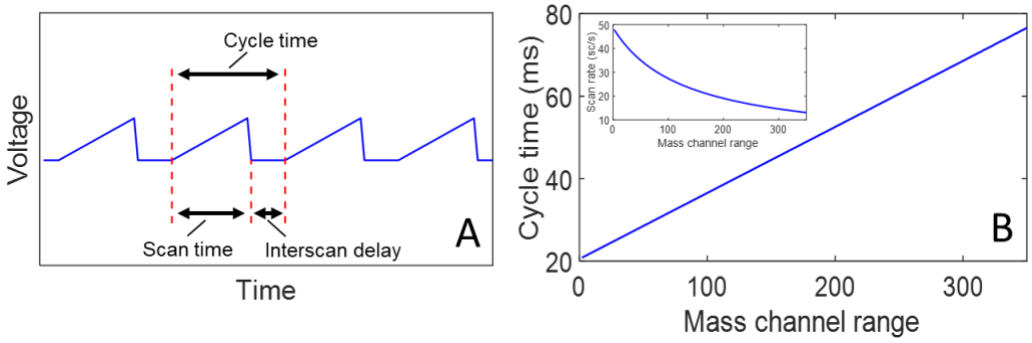


Figure 3.2. (A) Plot of scanned voltage versus time for the qMS in the context of Eqs. (1-3). The cycle time is defined as the time between starting two consecutive scans. The scan time is amount of the time that the voltages (DC and RF) are linearly ramped, and the interscan delay is the time the voltage is held constant between scans. (B) Relationship between the mass range and cycle time is shown for every mass range from 2 to 350 mass channels (m/z 45 to m/z 394) for the specific qMS instrument that was used. Insert shows relationship between scanned mass channel range and the scan rate, which is calculated by taking an inverse of cycle time.

For absolutely no visual distortions to appear in the GC×GC chromatogram, the cycle time defined in Eq. (1) must be divided evenly into the P_M. This can be achieved by adjusting either the P_M and/or the cycle time. For example, if a cycle time of 42 ms is applied, then the P_M could be adjusted to 1.974 s (47 mass spectra per P_M), or if a P_M of 2.000 s was used a cycle time of 40 ms could be used (50 mass spectra per P_M). However, if the cycle time does not divide evenly into P_M, the apparent time that the separation shifts per modulation is given by,

$$t_{\text{shift,mod}}\ (\text{s}) = \left| nint\left(\frac{P_{\text{M}}\ (\text{s})}{\text{cycle time (s)}}\right) - \frac{P_{\text{M}}\ (\text{s})}{\text{cycle time (s)}} \right| \times \text{cycle time (s)} \qquad (4)$$

Where the argument of *nint* is rounded to the nearest integer. Then the total apparent shift for the entire GC×GC separation run time (t_{sep}) is the sum of the shift for each modulation for the total number of modulations per chromatogram, N_{mod}.

$$t_{\text{shift,total}}\ (\text{s}) = \frac{t_{\text{sep}}\ (\text{s})}{P_{\text{M}}(\text{s})} \times\ t_{\text{shift,mod}}(\text{s}) = N_{\text{mod}}\ \times\ t_{\text{shift,mod}}(\text{s}) \qquad (5)$$

This phenomenon is illustrated in Fig. 3.3. In Fig. 3.3. A, a simulated unfolded peak GC×GC peak obtained at 20.00 Hz with P_{M} = 2 s is shown. The folded GC×GC chromatogram is given in Fig. 3B. Because the cycle time is evenly divisible into the P_{M}, no distortions appear. In Fig. 3.3. C, the same peak is shown sampled at 21.28 Hz. Here, the cycle time does not divide evenly into the P_{M}. As a result, the GC×GC chromatogram (Fig. 3.3. C) appears distorted, with each subsequent ^{2}D peak shifting to a higher apparent $^2t_{\text{R}}$.

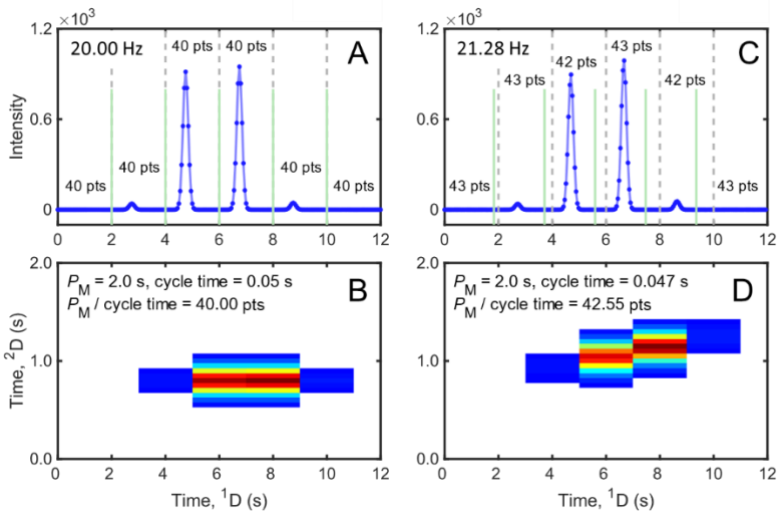


Figure 3.3. (A) Unfolded data for a simulated analyte at a scan rate of 20.00 Hz and a P_M of 2 s, which produces 40 data points (pts) per P_M. (B) Reshaped data for the analyte in (A). Here, the cycle time is divided evenly into the P_M. (C) Unfolded data for an analyte simulated at a scan rate of 21.28 Hz and P_M of 2 s, showing the variation of number of data points for every modulation (42-43 data points per P_M). (D) Reshaped chromatogram for the analyte in (C). Here, the cycle time does not divide evenly into the P_M, thus distorting the peak.

3.3 Experimental

All experiments were performed using a GC×GC-qMS instrument based upon an Agilent 5973 GC/qMS platform with a stock 7673B autoinjector (Agilent Technologies, Palo Alto, CA, USA), modified in-house with a high-speed, high-temperature six-port diaphragm valve, i.e., the HT-DV modulator (VICI model DV-12-1116 T, Valco Instruments Company Inc., Houston, TX, USA) fitted with 10 µL sample loop based on our previous work [20,31]. The duty cycle for this system when a P_M = 2 s is used was 13 %. This duty cycle was used in conjunction with a 1:20 split onto the ^{1}D column from the autoinjector, which results in an overall split of 1:150 of the sample. A schematic of the instrument is shown in Fig. 3.1. The HT-DV was operated using an in-house program written in LabVIEW (National Instruments, Austin, TX, USA). For all experiments, the inlet temperature was kept at 250 °C, the qMS ion source at 230 °C, and the mass filter at 150 °C. All data were collected after at least 2 hr of thermal equilibration for the ion source and filter. Ultra-high purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas that was additionally filtered using a triple gas filter (Restek Corporation, Bellefonte, PA, USA), designed to remove oxygen, moisture, and hydrocarbons.

Table 3.1. (gasoline and calibration MegaMix) and Table A.1. (JP8 fuel) provide method parameters for all sample analyses. A few details are explained briefly here. For the gasoline and calibration MegaMix analyses, a "less" polar PLOT-S (divinylbenzene vinylpyridine polymer) was used on ^{1}D with a 250 µm d_c. × 8 µm film thickness (Restek Corporation, Bellefonte, PA, USA) and a "more" polar Rtx-200 (crossbonded trifluoropropyl polysiloxane) on ^{2}D of 5 m length × 150 µm d_c. × 2 µm d_f (custom manufactured, Restek Corporation, Bellefonte, PA, USA). Since the PLOT-S stationary phase is relatively thick at 8 µm (even though the retention mechanism is nominally a surface interaction), the Rtx-200 stationary phase also needed to be a relatively thick

phase of 2 µm to balance the phase volume ratios of the two columns in order to effectively use the 2D separation space [32]. For the JP8 analysis to examine the quantitative reproducibility and *LOD* for the HT-DV modulate GC×GC-qMS instrument, a mid-polarity Rxi-17SilMS (crossbond diphenyl dimethyl polysiloxane) was used on ^{1}D with a 250 µm d_c × 0.25 µm film thickness (Restek Corporation, Bellefonte, PA, USA) and a non-polar Rtx-1 (crossbond dimethyl polysiloxane) on ^{2}D of 1.8 m length × 100 µm d_c × 0.1 µm d_f (custom manufactured, Restek Corporation, Bellefonte, PA, USA). Although a PLOT column was not used for ^{1}D for this part of the overall study, this was not deemed to be an issue since the quantitative reproducibility and *LOD* relate principally to the combination of using the HT-DV modulator and qMS detector.

Table 3.1. Experimental parameters for the studies performed.

Study	Scan rate Evaluation: Gasoline	Test mixture evaluation: 13 m PLOT-S	Test mixture evaluation: 5 m PLOT-S
^{1}D Column	5 m PLOT-S, 250 µm d_c × 8 µm d_f	13 m PLOT-S, 250 µm d_c × 8 µm d_f	5 m PLOT-S, 250 µm d_c × 8 µm d_f
^{2}D Column	5 m Rtx-200, 150 µm d_c × 2 µm d_f	5 m Rtx-200, 150 µm d_c × 2 µm d_f	5 m Rtx-200, 150 µm d_c × 2 µm d_f
Sample	Unleaded gasoline (locally sourced)	8260 Calibration MegaMix (Restek)	8260 Calibration MegaMix (Restek)
Injection volume	0.1 µl	0.1 µl	0.5 µl
Split ratio	20:1	20:1	20:1
P_M (s)	2 s	2 s	3 s
p_w (ms)	250 ms	250 ms	250 ms

Oven program	100 °C ramped to 250 °C at 15 °C/min and held for 5 min	100 °C ramped to 250 °C at 15 °C/min and held for 10 min	150 °C ramped to 250 °C at 10 °C/min and held for 20 min
Inlet pressure program, P_{inlet}	37.92 kPa (5.50 psig) ramped to 61.36 kPa (8.890 psig) at 2.34 kPa/min (0.34 psi/min)	83.01 kPa (12.04 psig) ramped to 126.86 kPa (18.40 psig) at 2.48 kPa/min (0.36 psi/min)	45.37 kPa (6.58 psig) ramped to 61.02 kPa (8.85 psig) at 1.59 kPa/min (0.23 psi/min)
Auxiliary pressure program, P_{aux}	191.81 kPa (27.82 psig) ramped to 288.20 kPa (8.890 psig) at 2.34 kPa/min (0.34 psi/min)	191.81 kPa (27.82 psig) ramped to 288.20 kPa (8.890 psig) at 9.65 kPa/min (1.40 psi/min)	224.77 kPa (32.6 psig) ramped to 288.20 kPa (41.80 psig) at 6.34 kPa/min (0.92 psi/min)

The HT-DV modulated GC×GC-qMS instrument is initially evaluated by analyzing a locally sourced gasoline sample. Subsequently, experimental, and computational methods to address the visual distortions when plotting GC×GC chromatograms with qMS data are discussed. Next, using the described methodology, the instrument is further evaluated by analyzing a complex test mixture (8260 MegaMix Calibration Mix, Restek Corporation, Bellefonte, PA, USA) and JP8 jet fuel. For the analysis of the calibration MegaMix, two 1D column lengths, 5 m and 13 m, are tested (Table 3.1.). When using the 5 m column, separations were performed at a higher temperature to investigate to what extent the column would provide sufficient retention using the shorter ^{1}D column length and higher temperature. Finally, to facilitate the separation of the heavier and more complex JP8 fuel sample, WCOT columns were used for both GC dimensions (as previously detailed) for the quantitative reproducibility and *LOD* studies. All the runs were performed in constant flow mode at 1.4 mL/min flow in the ^{1}D column and an average estimated

flow of 2.0 mL/min in the ^{2}D column, where the inlet pressure program and auxiliary pressure program information is provided in Table 3.1. and Table A.1.

Data processing was performed in MATLAB R2019a (The Mathworks, Inc., Natick, MA, USA). The data collected from the instrument is initially a 2D array, $m \times n$, where m is the number of spectra collected and j is the number of m/z scanned. To visualize a GC×GC chromatogram, the initial 2D array must be reshaped into a 3D data cube, $i \times j \times n$, where i is the number of data points per P_M and j is the number of modulations for the entire run. This is accomplished using the reshape function in MATLAB, which requires the analyst to enter integers for the size of each dimension of the new array. However, if the data were collected using a mass range that results in a cycle time that does not divide evenly into the P_M, the true value of i is not an integer, and if computational tools (resampling) were not used, then the value of i was rounded to the nearest integer. For example, given a cycle time of 49.505 ms (scan rate: 20.20 Hz) and P_M = 2 s, i is 40.40 data points (spectra) per P_M. Thus, i is rounded to down to 40 in order to apply the reshape function and visualize the GC×GC data. In this example, $t_{shift,mod}$ = 19.8 ms via Eq. (4). If the analyst desires to employ computational tools beyond what was just described, the computational method of resampling used for dealing with cycle times that do not divide evenly into the P_M is implemented herein, adapted from code published in Harvard Dataverse by Hantao and coworkers [30].

3.4 Results and Discussion

3.4.1 Initial Instrument Evaluation

A gasoline separation obtained using the HT-DV modulated GC×GC-qMS instrument and associated plots are shown in Fig. 3.4., using the PLOT-S column on ^{1}D coupled to a functionally more polar Rtx-200 column on ^{2}D. Thus, the appearance of the separation in Fig. 3.4. A is similar to that using a "standard" non-polar by polar GC×GC column configuration with WCOT columns

in the sense that saturated compounds are less retained on ^{2}D than the aromatic species. The ^{2}D peaks width-at-base (4σ) ranged from about 300 ms to 400 ms, so a mass range was selected that would simultaneously provide the fastest scan rate (shortest cycle time) possible concurrent with an appropriate mass range to provide the full mass spectrum for all components in the gasoline sample. Gasoline is a relatively light sample, so the mass range from 45 *m/z* to 228 *m/z* was sufficient to properly detect all components of the sample, and further using this range resulted in a relatively fast scan rate of 20.01 Hz, which is sufficiently close to a scan rate of 20.00 Hz (which is evenly divisible into the P_M of 2 s) so as not to visually distort the GC×GC chromatogram, as

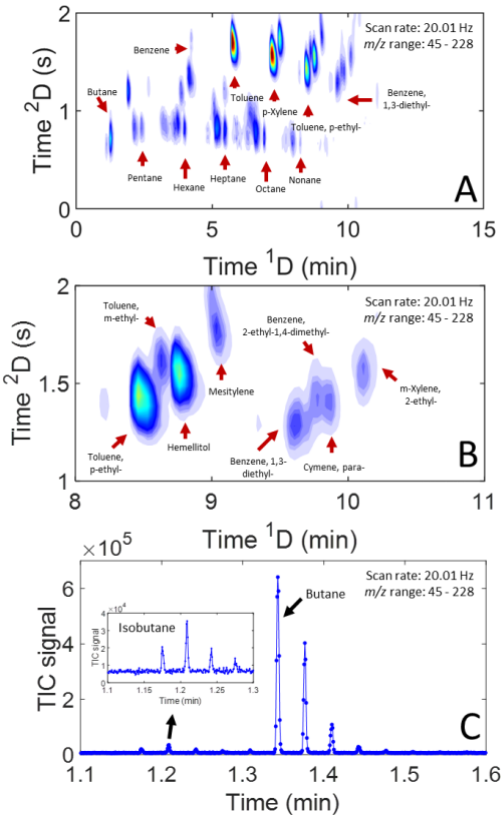


Figure 3.4. (A) Total ion current (TIC) chromatogram of the HT-DV modulated GC×GC-qMS separation of gasoline, where several alkanes (C4-C9) and some aromatics are identified. Isobutane and butane peaks were purposely plotted on a 10-fold more sensitive signal scale, to reveal these less concentrated peaks. (B) Zoom-in between 8 and 11 min of (A), showing various aromatic compounds readily separated using the PLOT-S × Rtx-200 column configuration. (C) The two first peaks separated in gasoline, isobutane and butane, are shown in an unfolded raw-data chromatogram, where circles on the trace show the positions of the mass spectrum scans (data points).

discussed in the next section. Here, each data point is plotted, demonstrating the sampling of the narrow 2D peaks by the qMS at 20.01 Hz (~50 ms cycle time resulting in ~6 to 8 spectra per ^{2}D peak width-at-base, 4σ). While 14.00 Hz scan rate would provide a broader mass range from 45 to 362 *m/z* (which was not needed) and a cycle time that is evenly divisible into the P_M, the peaks would be more sparsely scanned across their ^{2}D peak widths resulting in ~4 to 5 spectra per ^{2}D peak at width-at-base, 4σ. A zoom in view of the section between 8-11 min on ^{1}D and 1-2 s on ^{2}D in Fig. 4B highlights the separation of the aromatic species. The first two analytes observed in the gasoline separation are isobutane and butane, which highlights how the use of a high temperature diaphragm valve in conjunction with a PLOT-S column on ^{1}D provides an excellent separation of light compounds (<C5), as shown in an unfolded raw-data chromatogram in Fig. 4C, whereby each data point (spectrum) measured is plotted as a dot. Here we can see that scanning from *m/z* 45 to *m/z* 228 at a rate of 20.01 Hz generates about 6-8 data points (spectra) across each ^{2}D peak, which is a sufficient sampling density for quantitative analysis [33,34].

3.4.2. Visualization Issues with GC×GC-qMS and Correction Method

Next, gasoline sample separations were used to investigate the scan frequency constraints when using a qMS as a detector. Gasoline was selected due to the structured pattern this sample generates in 2D space, enabling visual distortions to be easily discerned. GC×GC data is usually processed in computational platforms as grid-like arrays. When the data is collected such that the cycle time divides evenly into the P_M, the number of spectra fits neatly into the array. However, if the cycle time does not divide evenly into the P_M, the magnitude of the remainder dictates the magnitude of the visual distortion that is obtained, via Eqs. (4) and (5). To highlight the relationship between the mass range, scan rate, and the extent of visual distortion, Fig. 3.5. shows GC×GC

chromatograms, unfolded data for butane, and plots of t_{shift} at three scan rates. Figure 3.5. A-C shows data for gasoline collected at 20.01 Hz, as highlighted previously in Fig. 3.4. Even though

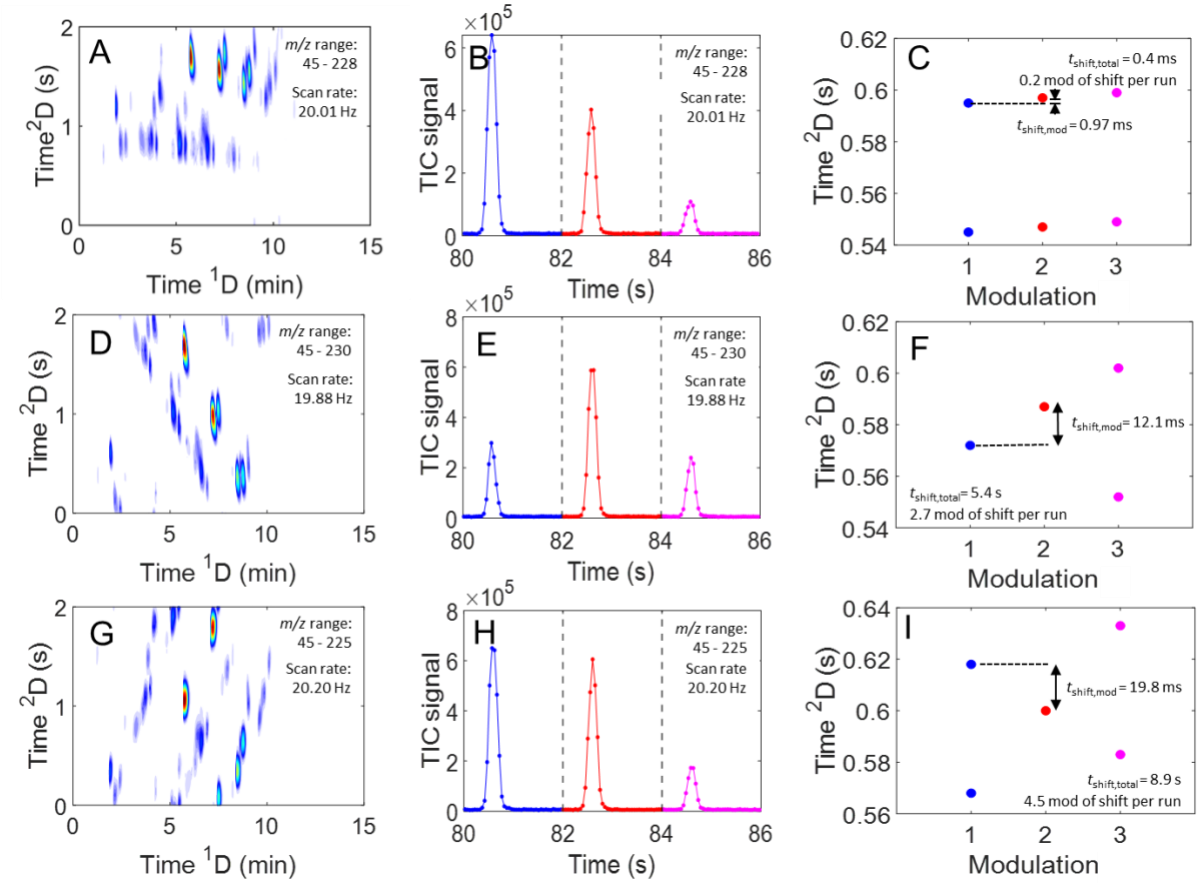


Figure 3.5. TIC chromatograms of the HT-DV modulated GC×GC-qMS separation of gasoline when different scan rates are used, and data is simply folded into the 2D space using an integer number of points in the second dimension (40 points for 2 s P_M), showing the shift of peaks. (A) GC×GC chromatogram collected at 20.01 Hz. (B) Inset of unfolded butane peak at 20.01 Hz. The first modulation is shown in blue, the second in red, and the third in pink. (C) Illustration of the amount of $t_{shift,mod}$ at 20.01 Hz. Each colored dot shows the ^2t_r for each ^{2}D peak and the measured $t_{shift,mod}$ is illustrated. (D) GC×GC chromatogram collected at 19.88 Hz. (E) Inset of unfolded butane peak at 19.88 Hz. (F) Illustration of the amount of $t_{shift,mod}$ at 19.88 Hz. (G) GC×GC chromatogram collected at 20.20 Hz. (H) Inset of unfolded butane peak at 20.20 Hz. (I) Illustration of the amount of $t_{shift,mod}$ at 20.20 Hz.

the scan rate of 20.01 Hz is not evenly divisible into the P_M (whereas 20.00 Hz would be evenly divisible), the remainder is such a small difference that it does not "add up" to enough of a shift to adversely impact the visualization of the GC×GC chromatogram. As shown in Fig. 3.5. C, $t_{shift,mod}$ is only 0.97 ms via Eq. (4), which adds up to a $t_{shift,total}$ of 405 ms via Eq. (5), equivalent to 0.21

modulations per 15 min run, thus 20.01 Hz does not create a significant visual distortion. However, in most cases, $t_{shift,total}$ can be high enough to create severe visual distortions. For example, Figs. 3.5. D-F show the sample collected at a scan rate of 19.88 Hz, the result of selecting mass range of *m/z* 45-300. In this example, $t_{shift,total}$ was 5.4 s (Fig. 3.5. H). Likewise, collecting data at *m/z* 45-225 generated a scan rate of 20.20 Hz, which also created severe visual distortions, shown in Fig. 3.5. G-I. Here, $t_{shift,mod}$ was similar at 19.8 ms with a $t_{shift,total}$ of 8.9 s.

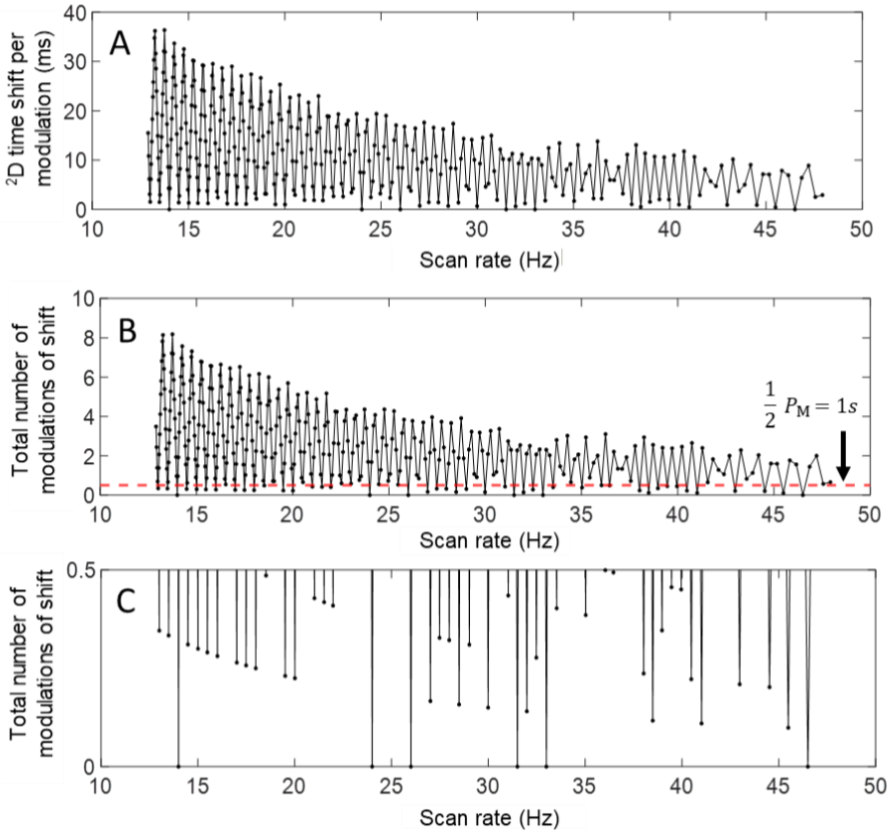


Figure 3.6. Plots showing errors associated with different scan rates using the qMS instrument. (A) The $t_{shift,mod}$ is calculated using Eq. (4) for every scan rate, produced when using mass ranges from 2 to 350 mass channels (m/z 45 to m/z 394). (B) The $t_{shift,total}$ is calculated in terms of how many modulations of shift using Eq. (5),using P_M = 2 s. The red line indicates the ½P_M of shift per 15 min run. (C) At this threshold the zoom-in of the graph is provided. A total of 44 scan rates fall under this threshold, out of which 6 have a shift of 0. These 44 scan rates are listed in Table 2.

Next, $t_{shift,mod}$ via Eq. (4) for each scan rate produced by the instrument was calculated by changing the mass range acquired, with the results presented in Fig. 3.6. A. An oscillating pattern

is observed in the plot, that dampens in range of error as the scan rate increases when a smaller mass range is being scanned. The $t_{shift,mod}$ obtained varies from 0 ms all the way up to almost 40 ms. To have the impact of the $t_{shift,mod}$ reflected in the context of the separation run time, Eq. (5) is applied, in this case using a P_M of 2 s and a run time of 15 min. The $t_{shift,total}$, in the units of number of modulations for the 15 min run is plotted against scan rate in Fig. 3.6. B. The upper limit of the shift error is nearly 8 modulations, while only 6 scan rates provide a shift error of exactly zero. A zoom-in version of Fig. 3.6. B to examine the scan rates that produce tshift,total ≤ ½PM, is shown in Fig. 3.6. C. An amount of distortion when tshift,total ≤ ½PM, was empirically deemed for the purpose of this study as a reasonably tolerable level. Overall, 44 scan rates over the 350 mass ranges were below this threshold, with 6 scan rates having a shift of zero (Table 3.2).

Table 3.2. Table of scan rates that produce less than half of the PM of shift, along with the scan rate, the mass range that corresponds to the scan rate and the number of modulations of shift in a 15 min GC×GC-qMS run calculated using Eq. (5) and converted to units of number of modulations.

Hz	ms/scan	Mass range	# modulations of shift
13.01	76.86	352	0.35
13.49	74.13	335	0.33
14.00	**71.43**	**318**	**0**
14.49	69.01	303	0.31
15.01	66.62	288	0.30
15.49	64.56	275	0.29
16.01	62.46	262	0.28
17.01	58.79	239	0.26
17.49	57.18	229	0.26
17.99	55.59	219	0.25
18.52	54.00	209	0.49
19.51	51.26	192	0.23
20.01	49.98	184	0.22
21.02	47.57	169	0.43
21.52	46.47	162	0.42
21.98	45.5	156	0.41
24.00	**41.67**	**132**	**0**

26.00	**38.46**	**112**	**0**
27.01	37.02	103	0.17
27.48	36.39	99	0.33
27.98	35.74	95	0.32
28.49	35.10	91	0.16
29.02	34.46	87	0.31
29.99	33.34	80	0.15
31.03	32.23	73	0.44
31.50	**31.75**	**70**	**0**
31.99	31.26	67	0.14
32.48	30.79	64	0.28
33.00	**30.30**	**61**	**0**
33.53	29.82	58	0.40
35.03	28.55	50	0.39
36.04	27.75	45	0.50
36.46	27.43	43	0.49
38.02	26.30	36	0.24
38.49	25.98	34	0.12
38.97	25.66	32	0.35
39.46	25.34	30	0.46
39.96	25.03	28	0.45
40.48	24.70	26	0.22
41.01	24.38	24	0.11
42.98	23.27	17	0.21
44.52	22.46	12	0.20
45.49	21.98	9	0.10
46.50	**21.51**	**6**	**0**

For laboratory situations in which implementing computational methods is not a priority, choosing a suitable scan rate manually using the method just discussed can help considerably to minimize the visual distortion when using the qMS detector with GC×GC. However, when a specific mass channel range is needed for a particular application, and that specific mass range with the qMS would result in significant $t_{shift,total}$, then computational methods can be used to address this situation. Recently, resampling the data was introduced in publicly available code to

deal with the issue by Hantao and co-workers [30]. In this approach, the data is resampled such that the cycle time would divide evenly into the P_M. This method was applied to the data that originally exhibited 4.5 modulations of shift in the 15 min run (Fig. 3.7. A), using scan rate of 20.20 Hz, when *m/z* from 45 to 225 were scanned, with the result presented in Fig. 3.7. B-D. The resampling method, as the name suggests, resamples data collected at rate of 20.20 Hz to a new fixed rate of 20.00 Hz, which is evenly divisible into the P_M, and when reshaped does not exhibit the wraparound in the folded GC×GC chromatogram (Fig. 3.7. B). The data was reregistered, such that alkanes are at the bottom of the chromatogram and aromatics are at the top of the chromatogram. The resampled and reregistered chromatogram is given in Fig. 3.7. C. Further, for peak width inspection, we examine three analytes: 3,5-dimethyloctane, 4-methyloctane and nonane (Fig. 3.7. D), which emphasize the high quality of the peaks following resampling.

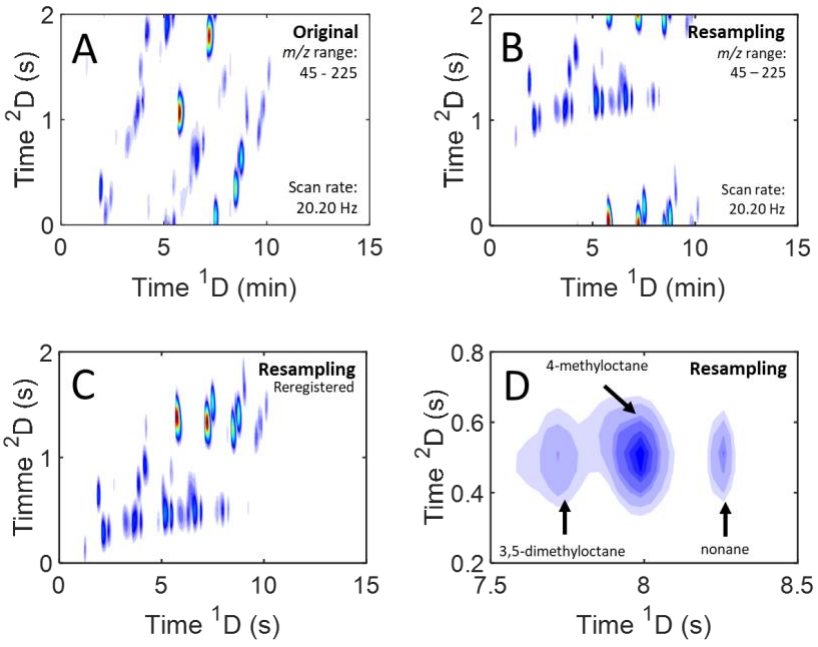


Figure 3.7. Computational resampling method solution to using a cycle time that does not divide evenly into the P_M. (A) Raw data collected at 20.20 Hz. (B) Literature method of resampling the vector of data into the closest integer scan rate. (C) Reregistered ^{2}D chromatogram to correct wraparound of aromatic compounds. (D) Zoom-in of a section between 7.5 and 8.5 min for the resampled and reregistered data, highlights that resampling provides non-distorted 2D peak shapes.

3.4.3. Final Instrument Evaluation

As part of this investigation, we briefly examined the separation of the Calibration MegaMix with a wide boiling point range of analytes, from 39.6 °C to 219 °C, thus containing a relatively broad range of analyte volatility from highly volatile to moderate volatility. Data were collected at the frequency of 20.01 Hz (m/z range: 45 - 228), and it was folded into a GC×GC chromatogram. The calibration MegaMix was run using two PLOT-S columns of different lengths

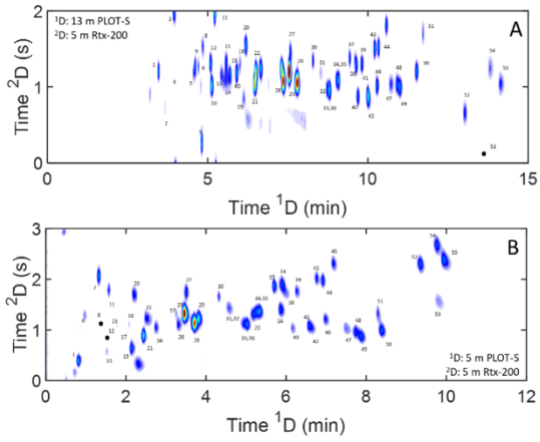


Figure 3.8. Separations of Calibration MegaMix using HT-DV modulated GC×GC-qMS with (A) 13 m and (B) 5 m (B) PLOT-S columns. Experimental details for both runs are provided in Table 1. For both separations, the beginning of each chromatogram was plotted on a more sensitive scale to better show less concentrated compounds. All analytes that were identified with numbers are listed in Table 3.3. For analytes 8, 12 and 53, that had very small signal compared to the other peaks, a dot indicates their peak location.

on ^{1}D starting at a lower temperature with the longer (13 m) column, anticipating obtaining the best separation (Fig. 3.8. A). Indeed, most of the analytes in the mix are readily separated and identified (Table 3.3.).

Table 3.3. Identified compounds and their boiling points in the Calibration MegaMix, with the peaks labeled by identification (ID) number in Fig. 3.8. Retention time entries of "N/A" indicate that analyte was not observed.

			13 m PLOT-S		**5 m PLOT-S**	
ID number	**Compound name**	**Boiling point, °C**	**$^{1}t_R$, min**	**$^{2}t_R$ apparent, s**	**$^{1}t_R$, min**	**$^{2}t_R$ apparent, s**
1	methylene chloride	39.6	3.468	1.231	0.814	0.407
2	diethyl ether	34.6	3.670	0.718	N/A	N/A

3	acrylonitrile	77.0	3.973	1.949	0.966	1.271
4	trans-1,2-dichloroethene	47.0	3.939	1.077	N/A	N/A
5	1,1-dichloroethane	57.2	4.545	1.179	N/A	N/A
6	bromochloromethane	68.1	4.714	1.282	N/A	N/A
7	propionitrile	97.0	4.815	0.256	1.322	2.085
8	tetrahydrofuran	66.0	4.848	1.538	1.373	1.220
9	1,1-dichloroethene	32.0	4.613	1.282	N/A	N/A
10	trichloromethane	61.2	5.118	1.026	1.525	0.610
11	methacrylonitrile	90.0	5.219	1.949	1.576	1.780
12	methylacrelate	80.0	5.084	1.333	1.475	0.966
13	2,2-dichloropropane	69.0	5.421	1.128	N/A	N/A
14	1,2,3-trichloropropane	156.0	5.589	1.128	5.898	1.932
15	1,2-dichloroethane	83.5	5.556	1.333	1.780	1.068
16	1,1,1-trichloroethane	74.0	5.690	1.179	N/A	N/A
17	trichloroethylyne	87.2	N/A	N/A	2.034	0.966
18	dibromomethane	97.0	5.993	1.282	N/A	N/A
19	isobutyl alcohol	108.0	6.128	0.821	2.136	0.661
20	1,4-dioxane	101.0	6.195	1.538	2.237	1.678
21	bromodichloromethane	90.0	6.465	1.026	2.441	0.864
22	methyl methacrylate	101.0	6.498	1.179	2.542	1.220
23	bromoform	149.1	N/A	N/A	5.186	1.322
24	1,1,2,2-tetrachloroethane	146.7	N/A	N/A	5.847	1.424
25	1,1,2-trichloroethane	113.9	7.542	1.179	3.458	1.322
26	toluene	110.6	7.374	1.077	3.305	1.119
27	1,3-dichloropropane	120.4	7.576	1.436	3.508	1.780
28	dibromochloromethane	119.0	7.811	1.077	3.712	1.119
29	ethyl methacrylate	117.0	N/A	N/A	3.814	1.220
30	chlorobenzene	132.0	8.283	1.333	4.322	1.678
31	1,1,1,2-tetrachloroethane	130.5	8.519	1.179	4.576	1.424
32	ethylbenzene	136.0	8.721	0.974	4.576	1.424
33	p-xylene	138.4	8.744	0.974	4.932	1.119
34	o-xylene	144.0	9.057	1.077	5.339	1.373
35	styrene	145.0	9.057	1.077	5.339	1.373
36	m-xylene	139.0	8.744	0.974	5.034	1.119
37	bromobenzene	156.0	9.428	1.385	5.695	1.881
38	4-chlorotoluene	162.0	9.630	1.282	6.000	1.729
39	2-chlorotoluene	159.2	9.832	1.282	6.254	1.780
40	cumene	152.4	9.697	0.872	6.153	1.068
41	n-propylbenzene	159.0	10.000	0.872	6.559	1.068
42	1,2,4-trimethylbenzene	169.0	10.030	0.821	6.661	1.017
43	1,4-dichlorobenzene	174.0	10.200	1.487	6.763	2.034
44	1,2-dichlorobenzene	180.5	10.340	1.538	6.915	1.983
45	1,3-dichlorobenzene	173.0	N/A	N/A	7.169	2.339
46	1,3,5-trimethylbenzene	164.7	10.270	1.026	6.966	1.220
47	tert-butylbenzene	169.0	10.740	1.026	7.525	1.017
48	sec-butylbenzene	173.0	10.910	1.026	7.729	0.966
49	p-cymene	177.0	11.010	0.974	7.881	0.864
50	n-butylbenzene	183.3	11.520	1.231	8.390	0.966
51	nitrobenzene	210.9	11.720	1.692	8.288	1.322

52	1,2,4-trichlorobenzene	214.4	13.030	0.667	9.356	2.288
53	hexachloro-1,3-butadiene	215.0	13.700	0.154	9.814	1.576
54	1,2,3-trichlorobenzene	218.5	13.800	1.282	9.763	2.695
55	naphthalene	218.0	14.140	1.026	9.966	2.390
56	2-chloroethanol	128.6	N/A	N/A	2.746	1.068
57	tetrachloroethene	121.1	N/A	N/A	3.254	1.322

Even the least retained analyte (methylene chloride) has a ^{1}D retention time of 3.5 min which means that using the longer PLOT column allows for potentially having even lighter analytes in the mix to be separated earlier in the run, e.g., as with the gasoline sample in Fig. 3.4. However, we wanted to explore what ^{2}D retention differences would be observed if the ^{1}D PLOT-S column was ~3-fold shorter (5 m) coupled with a starting temperature 50 °C higher, which is shown in Fig. 3.8. B. Surprisingly, the higher temperature and shorter column provides a very effective separation for the calibration MegaMix, and spreads out the analytes comfortably along the ^{1}D dimension with the first analyte (methylene chloride) now having a ^{1}D retention time of ~ 1 min, roughly 3.5-fold shorter than when 13 m column was used. However, most of the analytes have been separated essentially just as well on the shorter ^{1}D column, as on the longer ^{1}D column, for example, tert-butylbenzene (47) and sec-butylbenzene (48), which only minorly differ structurally and have very close boiling points of 169 °C and 174 °C, respectively. These examples demonstrate the exceptional ability of the PLOT-S column to retain and separate light compounds in GC×GC separations while using a relatively short ^{1}D column with the appropriately paired WCOT ^{2}D column, i.e., in this case the 5 m Rtx-200 with a 150 µm i.d. and a 2 µm film thickness.

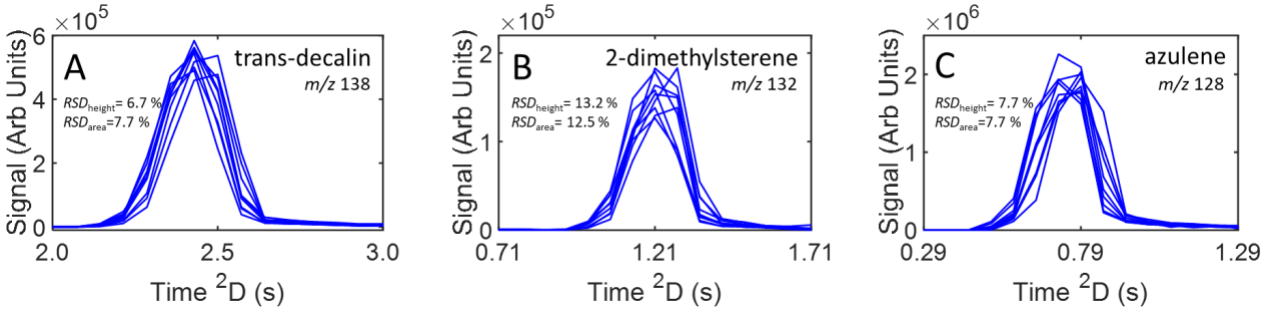


Figure 3.9. ^{2}D summed peak profiles used to evaluate peak height and area reproducibility for three analytes in the JP8 fuel using m/z indicated showing overlay of 9 injection replicates. (A) Trans-decalin (m/z 138) with a peak height RSD of 6.7 % and area RSD of 7.7 %. (B) 2-dimethylsterene (m/z 132) with a peak height RSD of 13.2 % and area RSD of 12.5 %. (C) Azulene (m/z 128) with a peak height RSD of 7.7 % and area RSD of 7.7 %.

Quantitative reproducibility and *LOD* figures-of-merit for the HT-DV modulated GC×GC-qMS instrument were investigated using a JP8 jet fuel separation (Fig. A.1). Reproducibility data in Fig. 3.9. uses the summed ^{2}D profiles at the most sensitive and selective *m/z* of three analytes native to the JP8: trans-decalin (*m/z* 138), 2-dimethylsterene (*m/z* 132) and azulene (*m/z* 128). The relative standard deviation (*RSD*) of these analyte peak heights and areas from 9 replicate runs of JP8 fuel were obtained: trans-decalin (Fig. 3.9. A) has RSD_{height} of 6.7 % and RSD_{area} of 7.7 %, while 2-dimethylstyrene (Fig. 9B) has RSD_{height} of 13.2 % and RSD_{area} of 12.5 %, and azulene (Fig. 9C) has RSD_{height} of 7.7 % and RSD_{area} of 7.7 %. We also spiked the JP8 fuel with three non-native analytes at 50 mg/L to determine *LOD*s. Summed ^{2}D profiles used to extrapolate to the *LOD* using the most sensitive and selective *m/z* for 1-chlorohexane (*m/z* 93), α-pinene (*m/z* 136) and bromobenzene (*m/z* 158) are provided in Fig. 3.10., for three injection replicates.

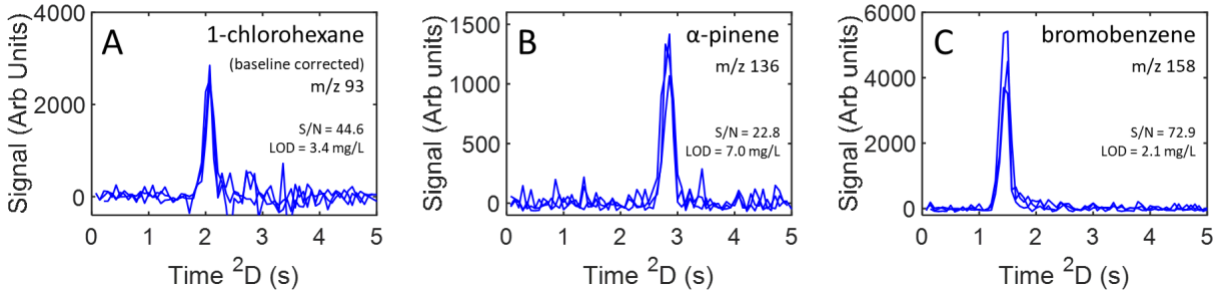


Figure 3.10. 2D summed peak profiles used to evaluate the peak height based LOD for three non-native analytes spiked at 50 mg/L in the JP8 fuel using m/z indicated showing overlay of 3 injection replicates. (A) 1-chlorohexane (m/z 93) has a LOD of 3.4 mg/L. (B) α-pinene (m/z 136) has a LOD of 7.0 mg/L. (C) Bromobenzene (m/z 158) has a LOD of 2.1 mg/L.

For 1-chlorohexane (shown after background correction in Fig. 3.10. A) was determined to have a *LOD* of 3.4 mg/L, α-pinene (Fig. 10B) was determined to have a *LOD* of 7.0 mg/L, and bromobenzene (Fig. 3.10. C) was determined to have a *LOD* of 2.1 mg/L. These results

demonstrate the ability of the HT-DV modulated GC×GC-qMS instrument to produce reproducible peaks with low *LOD*s, which makes this instrument appealing for lower-cost quantitative analysis applications.

3.5 Conclusion

Here we present the development and evaluation of a GC×GC-qMS instrument using a high-temperature diaphragm valve (HT-DV) modulator that can operate up to 325 °C. We show the impact to data reshaping when the cycle time does not divide evenly into the P_M, and implement a computational method, referred here as resampling, to correct this issue. This HT-DV modulated GC×GC-qMS instrument was also used to evaluate using a PLOT-S column as the ^{1}D column for gasoline analysis, with a relatively thick-film Rtx-200 column on ^{2}D, to balance the phase volume ratios of the columns. Also, the PLOT-S × Rtx-200 column combination was examined using two different PLOT-S column lengths coupled with a suitable temperature program in each case to separate the Calibration MegaMix that has a broad range of boiling point analytes from 39.6 °C to 219 °C. Using this HT-DV modulated GC×GC-qMS instrument, JP8 fuel was analyzed to evaluate quantitative reproducibility and *LOD*s. This instrument produced reproducible peak height (average RSD_{height} = 9.2 %, s = 3.5 %) and areas (average RSD_{area} = 9.3 %, s = 2.8 %) with low *LOD*s (average *LOD* = 4.2 mg/L, s = 2.5 mg/L). Overall, this research broadens the scope and potential for applying PLOT columns in GC×GC studies and in visualizing GC×GC-qMS data.

3.6 References

[1] Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, J. Chromatogr. Sci. 29 (1991) 227–231. https://doi.org/10.1093/chromsci/29.6.227.

[2] E.A. Higgins Keppler, H.L. Mead, B.M. Barker, H.D. Bean, Life Cycle Dominates the Volatilome Character of Dimorphic Fungus Coccidioides spp., MSphere 6 (2021). https://doi.org/10.1128/msphere.00040-21.

[3] A. Mani-Varnosfaderani, A. Gao, K.R. Poch, S.M. Caceres, J.A. Nick, J.E. Hill, Breath biomarkers associated with nontuberculosis mycobacteria disease status in persons with cystic fibrosis: a pilot study, J. Breath Res. 16 (2022). https://doi.org/10.1088/1752-7163/ac6bb6.
[4] P.H. Stefanuto, R. Romano, C.A. Rees, M. Nasir, L. Thakuria, A. Simon, A.K. Reed, N. Marczin, J.E. Hill, Volatile organic compound profiling to explore primary graft dysfunction after lung transplantation, Sci. Rep. 12 (2022) 2053. https://doi.org/10.1038/s41598-022-05994-2.
[5] D.M. Rasheed, A. Serag, Z.T. Abdel Shakour, M. Farag, Novel trends and applications of multidimensional chromatography in the analysis of food, cosmetics and medicine bearing essential oils, Talanta 223 (2021). https://doi.org/10.1016/j.talanta.2020.121710.
[6] L. Bai, J. Smuts, J. Schenk, J. Cochran, K.A. Schug, Comparison of GC-VUV, GC-FID, and comprehensive two-dimensional GC–MS for the characterization of weathered and unweathered diesel fuels, Fuel 214 (2018) 521–527. https://doi.org/10.1016/j.fuel.2017.11.053.
[7] J. Feldhausen, D.C. Bell, Z. Yang, C. Faulhaber, R. Boehm, J. Heyne, Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy, Fuel 326 (2022). https://doi.org/10.1016/j.fuel.2022.125002.
[8] K.A. Favela, M.J. Hartnett, J.A. Janssen, D.W. Vickers, A.J. Schaub, H.A. Spidle, K.S. Pickens, Nontargeted Analysis of Face Masks: Comparison of Manual Curation to Automated GCxGC Processing Tools, J. Am. Soc. Mass Spectrom. 32 (2021) 860–871. https://doi.org/10.1021/jasms.0c00318.
[9] J.E. Welke, K.P. Nicolli, K.C. Hernandes, A.C.T. Biasoto, C.A. Zini, Adaptation of an olfactometric system in a GC-FID in combination with GCxGC/MS to evaluate odor-active compounds of wine, Food Chem. 370 (2022). https://doi.org/10.1016/j.foodchem.2021.131004.
[10] Y. Zou, M. Gaida, F.A. Franchina, P.H. Stefanuto, J.F. Focant, Distinguishing between Decaffeinated and Regular Coffee by HS-SPME-GC×GC-TOFMS, Chemometrics, and Machine Learning, Molecules 27 (2022) 1806. https://doi.org/10.3390/molecules27061806.
[11] A.T. Lebedev, O. V. Polyakova, V.B. Artaev, M.B. Mednikova, E.A. Anokhina, Comprehensive two-dimensional gas chromatography-high resolution mass spectrometry with complementary ionization methods in the study of 5000-year-old mummy, Rapid Commun. Mass Spectrom. 35 (2021). https://doi.org/10.1002/rcm.9058.
[12] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A 1383 (2015) 151–159. https://doi.org/10.1016/j.chroma.2015.01.031.
[13] M.S.S. Amaral, Y. Nolvachai, P.J. Marriott, Comprehensive two-dimensional gas chromatography advances in technology and applications: biennial update, Anal. Chem. 92 (2020) 85–104. https://doi.org/10.1021/acs.analchem.9b05412.
[14] D. Zanella, J. Focant, F.A. Franchina, 30th Anniversary of comprehensive two-dimensional gas chromatography: Latest advances, Anal. Sci. Adv. 2 (2021) 213–224. https://doi.org/doi.org/10.1002/ansa.202000142.
[15] H. Cai, S.D. Stearns, A comprehensive two-dimensional gas chromatography valve modulation method using hold-release primary column flow for long secondary separation time with 100% transfer, J. Chromatogr. A 1569 (2018) 200–211. https://doi.org/10.1016/j.chroma.2018.07.064.
[16] W.C. Liao, C.F. Ou-Yang, C.H. Wang, C.C. Chang, J.L. Wang, Two-dimensional gas chromatographic analysis of ambient light hydrocarbons, J. Chromatogr. A 1294 (2013) 122–129. https://doi.org/10.1016/j.chroma.2013.04.008.

[17] R.A. Shellie, H.J. Cortes, R. Gras, J. Luong, Planar microfluidic devices in flow modulated comprehensive two dimensional gas chromatography for challenging petrochemical applications, Anal. Methods 5 (2013) 6598–6604. https://doi.org/10.1039/c3ay41198b.
[18] R.C.M. De Nijs, Analysis of Light Hydrocarbons C,-C5 with Porous Layer Open Tubular Fused Silica Columns of Aluminum Oxide Part 1 : The Column, J. High Resolut. Chromargr. (1981) 612–615. https://doi.org/10.1002/jhrc.1240041203.
[19] L. Mikaliunaite, D.S. Bell, R.E. Synovec, Application of Porous Layer Open Tubular Columns: Beyond Permanent Gases, LC GC. (2022) 476–483.
[20] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, J. Chromatogr. A. 1424 (2015) 127–133. https://doi.org/10.1016/j.chroma.2015.10.098.
[21] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.
[22] N.E. Heshka, D.B. Hager, Measurement of H2S in Crude Oil and Crude Oil Headspace Using Multidimensional Gas Chromatography, Deans Switching and Sulfur-selective Detection, J. Vis. Exp. (2015) 53416. https://doi.org/10.3791/53416.
[23] J.F. Griffith, W.L. Winniford, K. Sun, R. Edam, J.C. Luong, A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1226 (2012) 116–123. https://doi.org/10.1016/j.chroma.2011.11.036.
[24] T.J. Trinklein, D. V. Gough, C.G. Warren, G.S. Ochoa, R.E. Synovec, Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1609 (2020) 460488. https://doi.org/10.1016/j.chroma.2019.460488.
[25] L. Mikaliunaite, P.E. Sudol, C.N. Cain, R.E. Synovec, Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection, J. Chromatogr. A 1652 (2021). https://doi.org/10.1016/j.chroma.2021.462358.
[26] G. Purcaro, P.Q. Tranchida, C. Ragonese, L. Conte, P. Dugo, G. Dugo, L. Mondello, Evaluation of a rapid-scanning quadrupole mass spectrometer in an apolar × ionic-liquid comprehensive two-dimensional gas chromatography system, Anal. Chem. 82 (2010) 8583–8590. https://doi.org/10.1021/ac101678r.
[27] P. Korytár, J. Parera, P.E.G. Leonards, J. De Boer, U.A.T. Brinkman, Quadrupole mass spectrometer operating in the electron-capture negative ion mode as detector for comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1067 (2005) 255–264. https://doi.org/10.1016/j.chroma.2004.11.045.
[28] R.A. Shellie, P.J. Marriott, Comprehensive two-dimensional gas chromatography-mass spectrometry analysis of Pelargonium graveolens essential oil using rapid scanning quadrupole mass spectrometry, Analyst 128 (2003) 879–883. https://doi.org/10.1039/b304371a.
[29] G.S. Frysinger, R.B. Gaines, Comprehensive two-dimensional gas chromatography with mass spectrometric detection (GC x GC/MS) applied to the analysis of petroleum, HRC CC J. High Resolut. Chromatogr. 22 (1999) 251–255. https://doi.org/10.1002/(SICI)1521-4168(19990501)22:5<251::AID-JHRC251>3.0.CO;2-V.
[30] L.W. Hantao, C.A. Teixeira, V.H.C. Ferreira, CDF2MAT Automated SCRIPT to import NETCDF files to MATLAB | RESAMPLING added to correct RESHAPE for non-integer MS acquisition rates in GCxGC-MS data, Harvard Dataverse, V4. https://doi.org/10.7910/DVN/WMTEMF

[31] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.
[32] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1536 (2018) 16–26. https://doi.org/10.1016/j.chroma.2017.07.018.
[33] K. Maštovská, J. Hajšlová, S.J. Lehotay, Ruggedness and other performance characteristics of low-pressure gas chromatography–mass spectrometry for the fast analysis of multiple pesticide residues in food crops, J. Chromatogr. A 1054 (2004) 335–349. https://doi.org/10.1016/j.chroma.2004.08.061.
[34] M. Adahchour, M. Brandt, H.-U. Baier, R.J.J. Vreuls, A.M. Batenburg, U.A.Th. Brinkman, Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: principles and applications, J. Chromatogr. A 1067 (2005) 245–254. https://doi.org/10.1016/j.chroma.2004.09.094.

Chapter 4. Performance of thin-film porous layer open tubular columns in one-dimensional and comprehensive two-dimensional gas chromatography-mass spectrometry systems

4.1 Introduction

Gas chromatography (GC) is an essential analysis technique to separate volatile and semi-volatile compounds. GC is used to investigate samples in many areas including food [1,2], forensics [3,4], and biological samples [5]. Most of these separations are performed using wall-coated open tubular (WCOT) columns, as these columns can efficiently separate a wide variety of boiling point analytes [6–8]. WCOT columns are the state-of-the-art way to perform separations, however, they do not have good retention for gases and other light analytes. For matrixes that involve separations of light compounds the porous layer open tubular (PLOT) columns are traditionally used [9–11]. These columns have a porous layer of particles that are adhered to the walls of the capillary, and due to that they have many limitations. One of them is the maximum temperature to which the column can be heated, which makes it hard to separate any heavier compounds using PLOT columns. Their inability to separate heavier compounds has made their use to be very limited and so there is a need to innovate PLOT column design so a larger analyte boiling point range can be separated with them.

In the last two decades, there have been some advances made in the field of PLOT columns. PLOT columns have grown in popularity to be used in liquid chromatography (LC) for various applications due to their high efficiency, high permeability, and large peak capacity [12,13]. These columns are usually not commercially available and are made in a lab using different porous layer materials, like polymer [14], porous silica [15,16], and metal oxide [17,18]. These columns usually utilize very thin film thicknesses (in the range of 300-1000 nm)

and similar trends are observed in LC that we observe in gas chromatography, that with film thickness increase, retention factors increase and efficiency slowly decreases [15,19]. PLOT columns have also been used for cryoadsorbtion, which was developed by the National Institute of Standards and Technology (NIST) [20–22]. This use of PLOT columns depends on the thick film PLOT column due to its focus on being used as an in-field sample collection device [23]. In gas chromatography, similarly to LC, most advancement in the PLOT column field has been by new phase column innovations that are not commercially available [24,25]. However, unlike in LC, there has been very little work that has been done to investigate the use of thinner film PLOT columns.

More work has been done recently with PLOT columns in comprehensive two-dimensional (2D) gas chromatography (GC×GC). GC×GC is known to perform better than one-dimensional GC in peak capacity and is used as an analysis tool for complex sample separations [26–28]. Mostly for GC×GC PLOT columns have been used as the column of choice for both dimensions [29–31]. While these setups work great for light sample separations, it would be hard to separate any heavier analytes with them. Further, it would be hard to integrate this design with a mass spectrometry detector due to particle shedding and flow limitations of the PLOT columns. Another way PLOT columns have been used recently is not as the column itself but as a pre-column to the first dimension [32]. In this case, a very short 43 cm piece of PLOT column is attached before a regular WCOT column that is used in the first dimension. This is one solution to using PLOT columns in conjunction with WCOT columns and using them for higher boiling point analyte separations. Another solution is to use a short PLOT column as a first-dimension column, and a relatively long WCOT column as a second-dimension column [33]. This requires

matching the commercially available PLOT column with a relatively thick film WCOT column, to ensure the best utilization of two-dimensional space.

Here we present a new exploration of the use of thin-film PLOT columns in one-dimensional and two-dimensional gas chromatography separations. We are comparing the commercially available 8 µm PLOT-S column to the 4 µm and 2 µm film-thickness PLOT-S columns. We show how the thinner columns allow for heavier compounds to be separated with a PLOT column and most importantly the thinner film columns are still very good at separating extremely light analytes, so there is no downside to using thinner film PLOT columns. By testing them in two-dimensional (GC×GC) mode we also demonstrate a better utilization of two-dimensional space compared to an 8 µm PLOT-S column.

4.2 Experimental

4.2.1 Instrument parameters

All experiments were performed using a GC-qMS instrument based upon an Agilent 5973 GC/qMS platform with a stock 7673B autoinjector (Agilent Technologies, Palo Alto, CA, USA). For all experiments, the inlet temperature was kept at 250 °C, the qMS ion source at 230 °C, and the mass filter at 150 °C. All data were collected after at least 2 hr of thermal equilibration for the ion source and filter. Ultra-high purity helium (Grade 5, 99.999%, Praxair, Seattle, WA, USA) was used as the carrier gas that was additionally filtered using a triple gas filter (Restek Corporation, Bellefonte, PA, USA), designed to remove oxygen, moisture, and hydrocarbons. To perform GC×GC separations the GC-qMS instrument was modified in-house with a high-speed, high-temperature six-port diaphragm valve, i.e., the HT-DV modulator (VICI model DV-12-1116 T, Valco Instruments Company Inc., Houston, TX, USA) fitted with 10 µL sample loop based on

our previous work [34,35]. The HT-DV was operated using an in-house program written in LabVIEW (National Instruments, Austin, TX, USA).

Table 4.1. Experimental parameters for the one-dimensional comparison studies performed between different film thickness PLOT columns.

Sample	**Calibration MegaMix**			**Refinery Mix**		
Column	30 m PLOT-S, 250 µm d_c × 8 µm d_f	30 m PLOT-S, 250 µm d_c × 4 µm d_f	30 m PLOT-S, 250 µm d_c × 2 µm d_f	30 m PLOT-S, 250 µm d_c × 8 µm d_f	30 m PLOT-S, 250 µm d_c × 4 µm d_f	30 m PLOT-S, 250 µm d_c × 2 µm d_f
Injection volume	0.1 µl			40 µl		
Split ratio	50:1			Splitless		
Flow rate	1.4 ml/min					
Oven program	30 °C held for 1 min and ramped to 250 °C at 10 °C/min and held for 10 min	30 °C held for 1 min and ramped to 250 °C at 10 °C/min and held for 5 min	30 °C held for 1 min and ramped to 250 °C at 10 °C/min and held for 3 min	30 °C held for 1 min ramped to 250 °C at 10 °C/min	30 °C held for 1 min ramped to 250 °C at 10 °C/min	30 °C held for 1 min ramped to 250 °C at 10 °C/min
Inlet pressure program, P_{inlet}	95.11 kPa (13.794 psig) ramped to 209.31 kPa (30.358 psig) at 5.19 kPa/min (0.75 psi/min)	82.34 kPa (11.942 psig) ramped to 189.11 kPa (27.429 psig) at 4.85 kPa/min (0.71 psi/min)	76.41 kPa (11.082 psig) ramped to 179.75 kPa (26.070 psig) at 4.70 kPa/min (0.68 psi/min)	95.11 kPa (13.794 psig) ramped to 209.31 kPa (30.358 psig) at 5.19 kPa/min (0.75 psi/min)	82.34 kPa (11.942 psig) ramped to 189.11 kPa (27.429 psig) at 4.85 kPa/min (0.71 psi/min)	76.41 kPa (11.082 psig) ramped to 179.75 kPa (26.070 psig) at 4.70 kPa/min (0.68 psi/min)
Mass channel range, m/z	40 – 334			2 – 100		
Chromatogram figure	Fig. 4.1. A	Fig. 4.1. B	Fig. 4.1. C	Fig. 4.2. A	Fig. 4.2. B	Fig. 4.2. C

Three different PLOT-S (divinylbenzene vinylpyridine polymer) columns were compared in this study: commercially available 250 µm d_c. × 8 µm film thickness column (Restek Corporation, Bellefonte, PA, USA), and two specially made columns, one of 250 µm d_c. × 4 µm film thickness (Restek Corporation, Bellefonte, PA, USA) and one of 250 µm d_c. × 2 µm film thickness (Restek Corporation, Bellefonte, PA, USA). In 1D GC separations, 30 m length columns were used for all the film thicknesses. Two sets of each film thickness column, made at different times, were tested to make sure that the results were reproducible for non-commercially available PLOT columns. Tables 4.1. and 4.2. summarize temperature programs and experimental conditions for one-dimensional separations. In GC×GC separations 5 m of each

PLOT column was used on ^{1}D and a "more" polar Rtx-200 (crossbonded trifluoropropyl polysiloxane) on ^{2}D of 5 m length × 150 µm d_c. × 2 µm d_f (custom manufactured, Restek Corporation, Bellefonte, PA, USA). Table 4.3. summarizes temperature programs and experimental conditions for two-dimensional separations. Since the PLOT-S stationary phase is relatively thick at 8 µm (even though the retention mechanism is nominally a surface interaction), the Rtx-200 stationary phase also needed to be a relatively thick phase of 2 µm to balance the phase volume ratios of the two columns to effectively use the 2D separation space [36].

Table 4.2. Experimental parameters for the one-dimensional comparison study performed between the PLOT column and WCOT column.

Sample	**Refinery Mix**	
Column	30 m PLOT-S, 250 µm d_c × 2 µm d_f	30 m Rxi-5ms, 250 µm d_c × 0.25 µm d_f
Injection volume	40 µl	
Split ratio	Splitless	
Flow rate	1.4 ml/min	
Oven program	30 °C held for 1 min ramped to 135 °C at 5 °C/min	30 °C held for 1 min ramped to 80 °C at 5 °C/min
Inlet pressure program, P_{inlet}	76.41 kPa (11.082 psig) ramped to 126.86 kPa (18.399 psig) at 2.40 kPa/min (0.35 psi/min)	71.46 kPa (10.364 psig) ramped to 95.11 kPa (13.794 psig) at 2.37 kPa/min (0.34 psi/min)
Mass channel range, *m*/*z*	2 – 100	
Chromatogram figure	Fig 3A	Fig 3B

Table 4.3. Experimental parameters for the two-dimensional comparison study performed between the different film thickness PLOT columns in the first dimension.

Sample	**Calibration MegaMix**		**Refinery Mix**		**Gasoline**	
^{1}D Column	5 m PLOT-S, 250 µm d_c × 8 µm d_f	5 m PLOT-S, 250 µm d_c × 2 µm d_f	5 m PLOT-S, 250 µm d_c × 8 µm d_f	5 m PLOT-S, 250 µm d_c × 2 µm d_f	5 m PLOT-S, 250 µm d_c × 8 µm d_f	5 m PLOT-S, 250 µm d_c × 2 µm d_f
^{2}D Column	5 m Rtx-200, 150 µm d_c × 2 µm d_f					
Injection volume	0.5 µl		50 µl		0.1 µl	
Split ratio	50:1		Splitless		50:1	
P_M (s) p_w (ms)	3 s 500 ms		1 s 500 ms		3 s 500 ms	
Flow rate, ^{1}D	1.4 ml/min					
Flow rate, ^{2}D	2.0 ml/min					
Oven program	40 °C held for 1 min and ramped to 250 °C at 10 °C/min		40 °C held for 1 min ramped to 150 °C at 10 °C/min		40 °C held for 1 min and ramped to 250 °C at 10 °C/min	
Inlet pressure program, P_{inlet}	29.28 kPa (4.247 psig) ramped to 61.00 kPa (8.847 psig) at 1.51 kPa/min (0.31 psi/min)	24.48 kPa (3.551 psig) ramped to 51.76 kPa (7.506 psig) at 1.30 kPa/min (0.18 psi/min)	29.28 kPa (4.247 psig) ramped to 45.40 kPa (6.585 psig) at 1.51 kPa/min (0.31 psi/min)	24.48 kPa (3.551 psig) ramped to 38.27 kPa (5.551 psig) at 1.30 kPa/min (0.18 psi/min)	29.28 kPa (4.247 psig) ramped to 61.00 kPa (8.847 psig) at 1.51 kPa/min (0.31 psi/min)	24.48 kPa (3.551 psig) ramped to 51.76 kPa (7.506 psig) at 1.30 kPa/min (0.18 psi/min)
Auxiliary pressure program, P_{aux}	151.70 kPa (22.002 psig) ramped to 288.04 kPa (41.777 psig) at 6.49 kPa/min (0.94 psi/min)		151.70 kPa (22.002 psig) ramped to 224.50 kPa (32.561 psig) at 6.49 kPa/min (0.94 psi/min)		151.70 kPa (22.002 psig) ramped to 288.04 kPa (41.777 psig) at 6.49 kPa/min (0.94 psi/min)	
Mass channel range, *m*/*z*	45 – 228		40 – 223		45 – 228	
Chromatogram figure	Fig 4A	Fig 4B	Fig 5A	Fig 5B	Fig 6A	Fig 6B

4.2.2 Samples

Three samples were used to show the performance of thin-film PLOT columns: Calibration MegaMix, Refinery Mix, and gasoline. Calibration MegaMix was an analyte mix with a wide range of boiling points, containing 75 analytes, used to see how the maximum temperature needed to get heavier analytes off the PLOT column changes when using thinner-film PLOT columns (8260 MegaMix Calibration Mix, Restek Corporation, Bellefonte, PA, USA). Refinery mix (DCG Refinery Gas Standard #2, Restek Corporation, Bellefonte, PA, USA) containing 27 analytes was used as an extra light analyte mixture to make sure that with the use of thinner-film PLOT columns retention is not lost on the isomers of gases. Locally sourced gasoline was used only in GC×GC separations as a real-world sample.

4.3 Results and Discussion

We have evaluated the performance of different film thickness PLOT-S columns: 8 µm, 4 µm, and 2 µm using different samples in GC-MS and GC×GC-MS systems. First, we want to compare one-dimensional gas separations (GC) of Calibration Megamix separated using 30 m PLOT-S columns of different films shown in Fig 4.1. A-C. Separation performed on a commercially available PLOT-S column of 8 µm (Fig 4.1. A) shows how even separating a relatively light mixture, a thicker film column needs to be at its maximum temperature of 250 °C to get propylbenzene to come off the column. Propylbenzene has a boiling point of 159 °C.

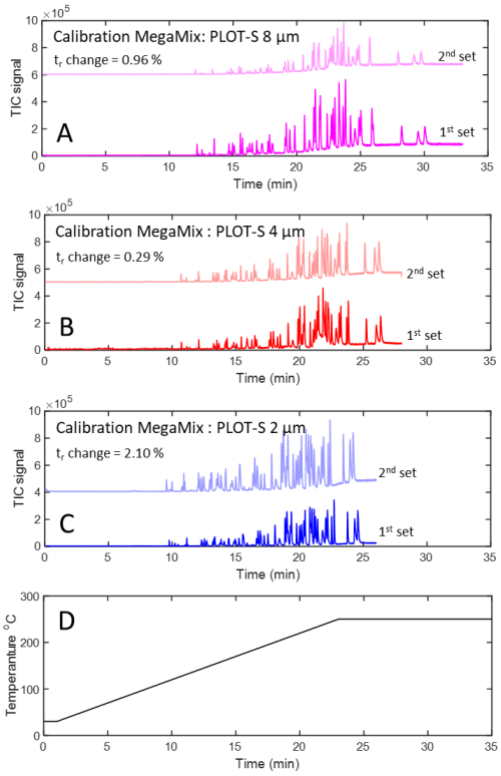


Figure 4.1. One-dimensional separations of Calibration MegaMix using (A) 8 µm film thickness column, (B) 4 µm film thickness column, and (C) 2 µm film thickness column. Chromatograms of different hues but the same color are replicate separations performed using the same thickness column that was manufactured at different times. (D) Temperature program for the separations shown in A-C.

Meanwhile, when we look at the same sample separation (Fig 4.1. B) on a thinner film PLOT-S column of 4 µm, the same compound of propylbenzene comes off the column at 237 °C (13 °C

cooler). This column with 4 µm film column needs to be at its maximum temperature of 250 °C to get sec-butylbenzene off the column. Sec-butylbenzene has a boiling point of 174 °C, which is 15 °C higher boiling point range, compared to the 8 µm film PLOT-S column. Finally, when we compare those with the separation on the thinnest PLOT-S column with 2 µm film (Fig 4.1. C), it seems propylbenzene comes off the column at 225 °C (25 °C cooler than the 8 µm film column), meanwhile, sec-butylbenzene comes off the column at 237 °C (13 °C cooler than 4 µm film column). This column with 2 µm film column needs to be at its maximum temperature of 250 °C to get 1,2,4-trichlorobenzene off the column. 1,2,4-trichlorobenzene has a boiling point of 214 °C, which is 55 °C higher boiling point range compared to the 8 µm film PLOT-S column.

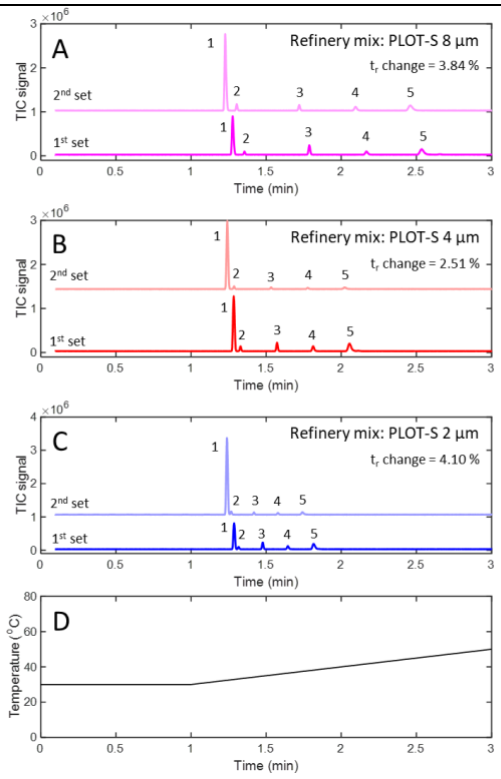


Figure 4.2. One-dimensional separations of the lightest 5 analytes in Refinery Mix using (A) 8 µm film thickness column, (B) 4 µm film thickness column, and (C) 2 µm film thickness column. The five analytes numbered are: 1 – nitrogen, 2 – methane, 3 – carbon dioxide, 4 – ethene, and 5 – ethane. Chromatograms of different hues but the same color are replicate separations performed using the same thickness column that was manufactured at different times. (D) Temperature program for the separations shown in A-C.

The other important aspect of exploring applications of thin-film PLOT columns is testing their ability to separate very light analytes that are generally in gaseous form. This was done by looking at the lightest part of the refinery mix separation using different film columns shown in Figure 4.2. A-C. It can be seen how when 8 µm film PLOT-S column is used, the first 5 analytes that separate are: 1 – nitrogen (b.p. -195.8 °C), 2 – methane (b.p. –161.5 °C), 3 – carbon dioxide (b.p. –78.5 °C), 4 – ethene (b.p. –103.7 °C), and 5 – ethane (b.p. –89.0 °C). This separation spans over 2.5 minutes on an 8 µm film PLOT-S column (Fig 4.2. A), however, it goes down to over 2 min when using a 4 µm film PLOT-S column (Fig. 4.2. B) and to about 1.7 min on 2 µm film PLOT-S column (Fig 4.2. B). However, the most remarkable part about the separation performed by the 2 µm film PLOT-S column, is that even know the separation time decreases significantly for these 5 analytes, the baseline resolution (R>1.5) is achieved for analytes 2-5, and the resolution between analyte 1 and 2 is about 1. This is an impressive improvement in separation possibilities of a thinner-film PLOT column over a commercially available thick-film column, as not only is this column able to separate higher boiling point analytes (by 55 °C) within the temperature program, but also it does not lose its selectivity for extremely light analytes. That can further be seen when exploring how a full separation of the Refinery mix on a 2 µm film PLOT-S column (Fig. 4.3. A) compares to the one performed on an Rxi-5 wall-coated column (Fig. 4.3. B). It can be seen how the thin-film PLOT column separates the analytes by their carbon count and separates isomers with the same carbon number into fully resolved peaks. Meanwhile, at the same separation temperature as the PLOT column, the Rxi-5 struggles to resolve analytes present in the Refinery mix, only barely separating them by the carbon count.

Further, as we are exploring non-commercially available columns, all the 1D runs were performed on two sets of columns (that were individually made at different times). We can see that for Calibration MegaMix separation on a commercially available 8 µm film column (Fig. 4.1. A), the retention time shift between two sets of columns for the same analyte was 0.96 %. But for the thinner film columns of 4 µm (Fig. 4.1. B) and 2 µm (Fig. 4.1. C) film thickness, the retention time shift between two sets of columns for the same analyte was 0.29 % and 2.10 %, respectively. Those are very comparable to the commercially available column, and on average (1.20 %) perform only slightly worse than such columns.

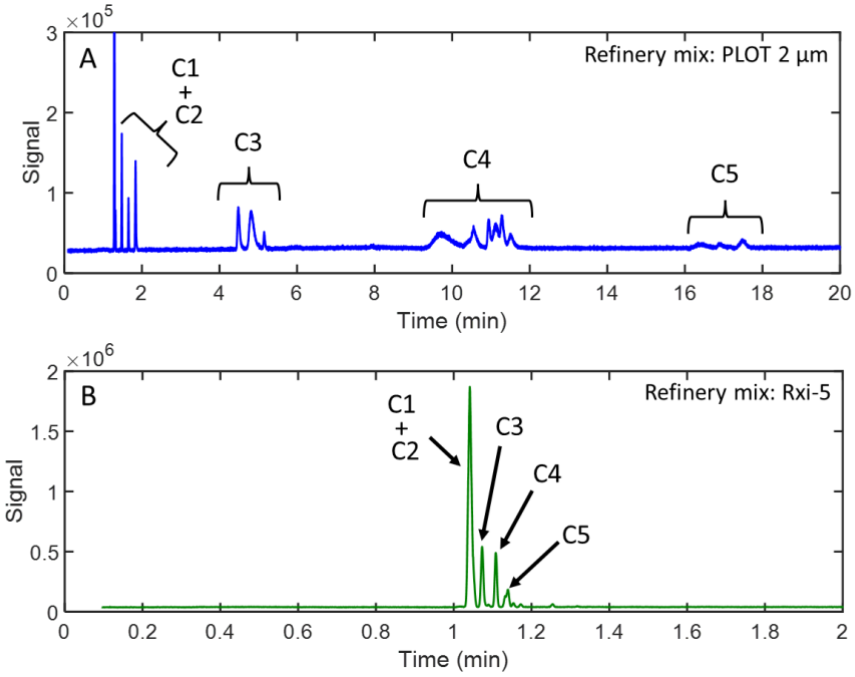


Figure 4.3. One-dimensional separations of Refinery Mix using (A) 2 µm film thickness PLOT-S column and (B) Rxi-5 wall coated column.

Now that we have shown how much better 1D separations are with thin-film PLOT-S column, we want to explore some of the GC×GC separations comparing 8 µm and 2 µm film columns used in the first dimension and their performance separating Calibration MegaMix, Refinery Mix, and a real-world sample of gasoline. First, looking at separations of the

Calibration MegaMix on 8 µm (Fig. 4.4. A) and 2 µm (Fig. 4.4. B) film columns we can note that the separation is generally shifted to the left on the 2 µm column compared to the one on 8 µm column. To look at more exact numbers, we can track analyte o-xylene on the temperature graph (Fig. 4.4. C) and see that it comes off the 8 µm column at 193 °C, while the same analyte comes off the 2 µm column at 172 °C (21 °C less).

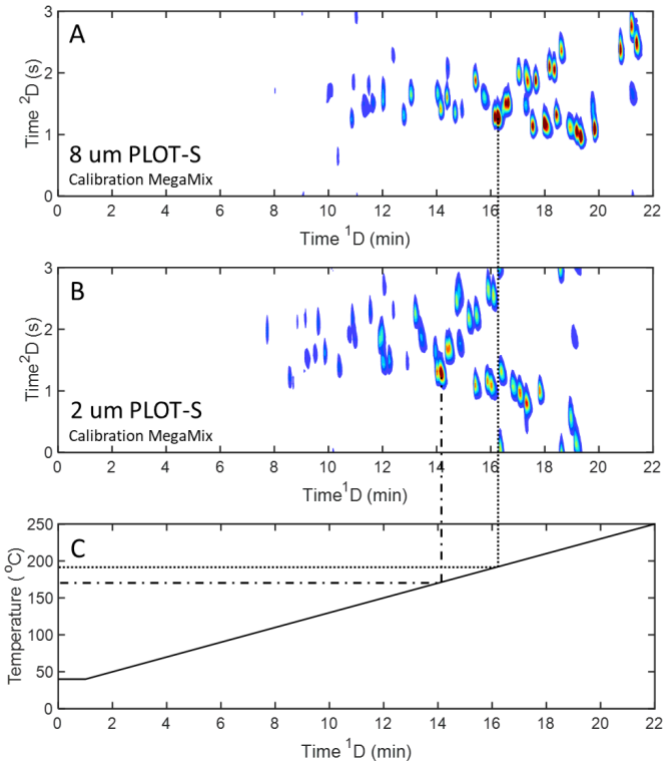


Figure 4.4. Two-dimensional separations of Calibration Megamix using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B.

Further, the same pattern can be seen when looking at Refinery mix separations shown in Figure 4.5. Here, we can track the analyte of isobutane on the temperature graph (Fig. 4.5. C) and see that it comes off the 8 µm column (Fig. 4.5. A) at 78 °C, while the same analyte comes off the 2 µm column (Fig 4.5. B) at 64 °C (14 °C less). Refinery mix separation on a thinner-film column shows again that it utilizes the separation space better, by not losing separation on the light components, but separating them at lower temperatures.

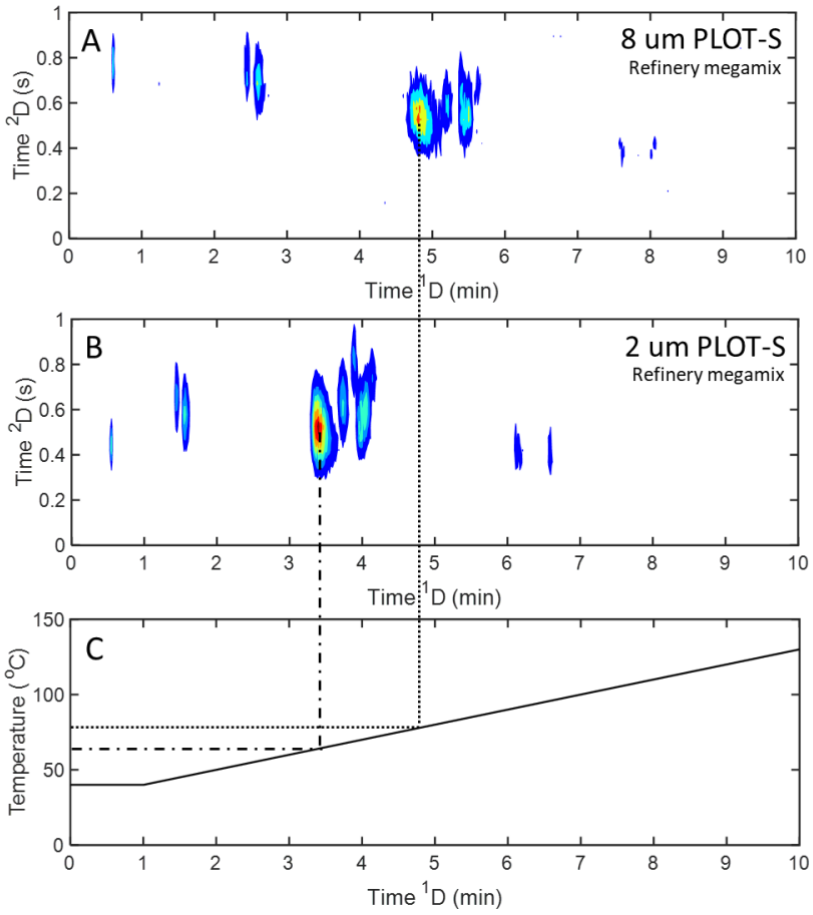


Figure 4.5. Two-dimensional separations of Refinery Mix using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B.

When looking at a real-world sample of gasoline in Figure 4.6., the same applies here as in the separations of Calibration MegaMix (Fig. 4.4.) and Refinery mix (Fig. 4.5.), where the analytes in the first dimension seem to be coming out colder. Here we can also track the analyte of 3-methylhexane on the temperature graph (Fig. 4.6. C) and see that it comes off the 8 µm column at 162 °C, while the same analyte comes off the 2 µm column at 142 °C (20 °C less). Further, these separations of gasoline show very clearly that the benefit of having analytes come out cooler is that it allows for much better utilization of 2D space. Analytes coming out sooner allows for better separation on the second-dimension column, which is essential when using PLOT columns in GC×GC separations.

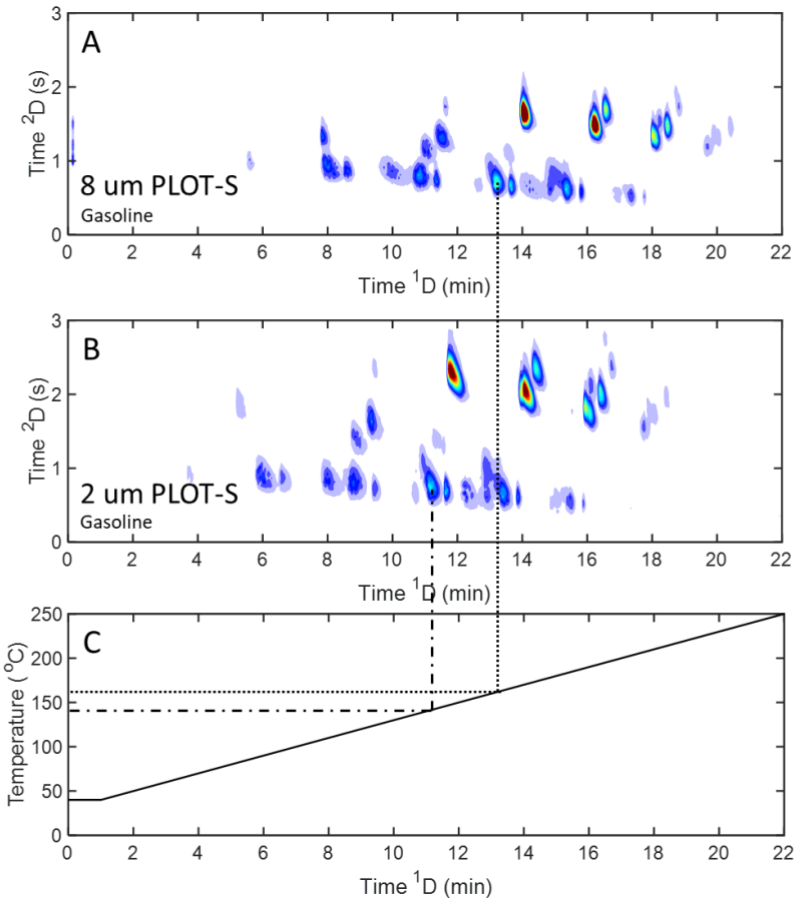


Figure 4.6. Two-dimensional separations of gasoline using (A) 8 µm film thickness PLOT-S column in the first dimension and (B) 2 µm film thickness PLOT-S column in the first dimension, but the same second dimension column. (C) Temperature program for the separations shown in A-B.

4.4. Conclusion

Here we present the performance evaluation of thin-film porous layer open tubular (PLOT) columns. We compare commercially available 8 µm film thickness PLOT-S columns to custom-made 4 µm and 2 µm film thickness columns. Both custom-made columns were tested in duplicate, with columns being made at different times. Comparing one-dimensional separations, we found that the column with 2 µm film column needs to be at its maximum temperature of 250 °C to get 1,2,4-trichlorobenzene off the column, compared to propylbenzene for the commercially available 8 µm film column. Propylbenzene has a boiling point of 159 °C and 1,2,4-trichlorobenzene has a boiling point of 214 °C. This shows how the 2 µm film column has

a higher boiling point range of 55 °C compared to the 8 µm film PLOT-S column. But also, through separations of refinery mix (very light mix) 2 µm film column is shown not to lose its selectivity for extremely light analytes. This was further shown by comparing a full separation of the Refinery mix on a 2 µm film PLOT-S column with the one performed on an Rxi-5 wall-coated column. The thin-film PLOT column separated the analytes by their carbon count and separated isomers with the same carbon number into fully resolved peaks while Rxi-5 struggled to resolve analytes present in the Refinery mix, only barely separating them by the carbon count. Further, we also explored some of the GC×GC separations comparing 8 µm and 2 µm film PLOT-S columns used in the first dimension and their performance separating Calibration MegaMix, Refinery Mix, and a real-world sample of gasoline. These separations on a thinner-film column showed that the column utilizes the separation space better compared to commercially available columns, by not losing separation on the light components, but separating them at lower temperatures. This work shows the possibility of using thinner-film PLOT columns for a wider range of samples that they are currently used for including heavier fuels and per- and poly-fluoroalkyl substances (PFAS).

4.5 References

[1] S. Wang, H. Chen, B. Sun, Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS), Food Chemistry 315 (2020) 126158. https://doi.org/10.1016/j.foodchem.2019.126158.
[2] O.A. Adebo, S.A. Oyeyinka, J.A. Adebiyi, X. Feng, J.D. Wilkin, Y.O. Kewuyemi, A.M. Abrahams, F. Tugizimana, Application of gas chromatography–mass spectrometry (GC-MS)-based metabolomics for the study of fermented cereal and legume foods: A review, International Journal of Food Science & Technology 56 (2021) 1514–1534. https://doi.org/10.1111/ijfs.14794.
[3] P.E. Sudol, K.M. Pierce, S.E. Prebihalo, K.J. Skogerboe, B.W. Wright, R.E. Synovec, Development of gas chromatographic pattern recognition and classification tools for compliance and forensic analyses of fuels: A review, Analytica Chimica Acta 1132 (2020) 157–186. https://doi.org/10.1016/j.aca.2020.07.027.
[4] C. Ribeiro, R. Gonçalves, M.E. Tiritan, Separation of Enantiomers Using Gas Chromatography: Application in Forensic Toxicology, Food and Environmental Analysis, Critical

Reviews in Analytical Chemistry 51 (2021) 787–811. https://doi.org/10.1080/10408347.2020.1777522.
[5] H.-H. Chiu, C.-H. Kuo, Gas chromatography-mass spectrometry-based analytical strategies for fatty acid analysis in biological samples, Journal of Food and Drug Analysis 28 (2020) 60–73. https://doi.org/10.1016/j.jfda.2019.10.003.
[6] S. Tuo, C. Liu, C. Wang, B. Kong, H. Lu, K. Zhong, Y. Li, W. Liu, J. Yu, Evaluation of Fourier deconvolution ion mobility spectrometer as high-performance gas chromatography detector for the analysis of plant extract flavors, Journal of Chromatography A 1714 (2024) 464560. https://doi.org/10.1016/j.chroma.2023.464560.
[7] G. Mishra, P. Gupta, A. Kumar, A. Purohit, V. Tak, D. Pardasani, Gas chromatography tandem mass spectrometric analysis of alkylphosphonofluoridic acids as verification targets of nerve agents, Journal of Chromatography A 1716 (2024) 464645. https://doi.org/10.1016/j.chroma.2024.464645.
[8] G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal Bioanal Chem 415 (2023) 2411–2423. https://doi.org/10.1007/s00216-022-04348-1.
[9] H. Yun, M.L. Lee, K.E. Markides, Charcoal porous layer open tubular column gas chromatography for permanent gas analysis, Journal of Microcolumn Separations 7 (1995) 207–212. https://doi.org/10.1002/mcs.1220070304.
[10] D.S. Bell, R. Synovec, L. Mikaliunaite, Application of Porous Layer Open Tubular Columns: Beyond Permanent Gases, LCGC North America 40 (2022) 476–483. https://www.chromatographyonline.com/view/application-of-porous-layer-open-tubular-columns-beyond-permanent-gases (accessed October 13, 2022).
[11] E. de Vanssay, P. Capilla, D. Coscia, L. Do, R. Sternberg, F. Raulin, Gas chromatography of Titan's atmosphere: IV. Analysis of permanent gases in the presence of hydrocarbons and nitriles with a Molsieve PLOT capillary column, Journal of Chromatography A 639 (1993) 255–259. https://doi.org/10.1016/0021-9673(93)80261-6.
[12] D.A. Collins, E.P. Nesterenko, B. Paull, Porous layer open tubular columns in capillary liquid chromatography, Analyst 139 (2014) 1292–1302. https://doi.org/10.1039/C3AN01869E.
[13] W. Zhou, W. Sun, Y. Liu, Z. Mao, Z. Chen, Ionic liquid-copolymerized monolith based porous layer open tubular column for CEC-MS analysis, Talanta 209 (2020) 120556. https://doi.org/10.1016/j.talanta.2019.120556.
[14] C. Kulsing, R. Knob, M. Macka, P. Junor, R.I. Boysen, M.T.W. Hearn, Molecular imprinted polymeric porous layers in open tubular capillaries for chiral separations, Journal of Chromatography A 1354 (2014) 85–91. https://doi.org/10.1016/j.chroma.2014.05.065.
[15] T. Hara, S. Futagami, S. Eeltink, W. De Malsche, G.V. Baron, G. Desmet, Very High Efficiency Porous Silica Layer Open-Tubular Capillary Columns Produced via in-Column Sol–Gel Processing, Anal. Chem. 88 (2016) 10158–10166. https://doi.org/10.1021/acs.analchem.6b02713.
[16] X. Liu, S. Sun, R. Nie, J. Ma, Q. Qu, L. Yang, Highly uniform porous silica layer open-tubular capillary columns produced via in-situ biphasic sol–Gel processing for open-tubular capillary electrochromatography, Journal of Chromatography A 1538 (2018) 86–93. https://doi.org/10.1016/j.chroma.2018.01.024.
[17] Y. Kimura, S. Shibasaki, K. Morisato, N. Ishizuka, H. Minakuchi, K. Nakanishi, M. Matsuo, T. Amachi, M. Ueda, K. Ueda, Microanalysis for MDR1 ATPase by high-performance

liquid chromatography with a titanium dioxide column, Analytical Biochemistry 326 (2004) 262–266. https://doi.org/10.1016/j.ab.2003.12.012.
[18] Y. Sekiguchi, N. Mitsuhashi, Y. Inoue, H. Yagisawa, T. Mimura, Analysis of sugar phosphates in plants by ion chromatography on a titanium dioxide column with pulsed amperometric detection, Journal of Chromatography A 1039 (2004) 71–76. https://doi.org/10.1016/j.chroma.2004.02.015.
[19] Q. Luo, T. Rejtar, S.-L. Wu, B.L. Karger, Hydrophilic interaction 10μm I.D. porous layer open tubular columns for ultratrace glycan analysis by liquid chromatography–mass spectrometry, Journal of Chromatography A 1216 (2009) 1223–1231. https://doi.org/10.1016/j.chroma.2008.09.105.
[20] T.J. Bruno, Simple, Quantitative Headspace Analysis by Cryoadsorption on a Short Alumina PLOT Column*, Journal of Chromatographic Science 47 (2009) 569–574. https://doi.org/10.1093/chromsci/47.7.569.
[21] J.E. Nichols, M.E. Harries, T.M. Lovestead, T.J. Bruno, Analysis of arson fire debris by low temperature dynamic headspace adsorption porous layer open tubular columns, Journal of Chromatography A 1334 (2014) 126–138. https://doi.org/10.1016/j.chroma.2014.01.080.
[22] T.M. Lovestead, T.J. Bruno, Trace Headspace Sampling for Quantitative Analysis of Explosives with Cryoadsorption on Short Alumina Porous Layer Open Tubular Columns, Anal. Chem. 82 (2010) 5621–5627. https://doi.org/10.1021/ac1005926.
[23] M.E. Harries, K.M. Jeerage, Preservation of vapor samples on adsorbent alumina capillaries and implications for field sampling, Journal of Chromatography A 1660 (2021) 462670. https://doi.org/10.1016/j.chroma.2021.462670.
[24] K.D. Otaif, A.-Y. Badjah-Hadj-Ahmed, Z.A. ALOthman, Preparation of UiO-66 MOF-Bonded Porous-Layer Open-Tubular Columns Using an In Situ Growth Approach for Gas Chromatography, Molecules 29 (2024) 2505. https://doi.org/10.3390/molecules29112505.
[25] Z. Jia, C. Du, L. Zhu, Z. Wang, Retention performance of alumina porous layer open-tubular column coated with γ–alumina nanoparticles in the highly volatile hydrocarbons separation, Journal of Chromatography A 1687 (2023) 463657. https://doi.org/10.1016/j.chroma.2022.463657.
[26] L. Mikaliunaite, R.E. Synovec, Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data, Journal of Chromatography A 1730 (2024) 465093. https://doi.org/10.1016/j.chroma.2024.465093.
[27] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, Journal of Chromatography A 1662 (2022) 462735. https://doi.org/10.1016/j.chroma.2021.462735.
[28] C.N. Cain, M. Gaida, P.-H. Stefanuto, J.-F. Focant, R.E. Synovec, S.C. Jackels, K.J. Skogerboe, Investigating sensory-classified roasted arabica coffee with GC × GC-TOFMS and chemometrics to understand potato taste defect, Microchemical Journal 196 (2024) 109578. https://doi.org/10.1016/j.microc.2023.109578.
[29] A.Y. Sholokhova, Y.V. Patrushev, V.N. Sidelnikov, A.K. Buryak, Analysis of light components in pyrolysis products by comprehensive two-dimensional gas chromatography with PLOT columns, Talanta 209 (2020) 120448. https://doi.org/10.1016/j.talanta.2019.120448.

[30] Y.V. Patrushev, V.N. Sidelnikov, Selection of the porous layer open tubular columns for separation of light components in comprehensive two-dimensional gas chromatography, Journal of Chromatography A 1579 (2018) 83–88. https://doi.org/10.1016/j.chroma.2018.10.015.
[31] L. Mikaliunaite, P.E. Sudol, C.N. Cain, R.E. Synovec, Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection, Journal of Chromatography A 1652 (2021) 462358. https://doi.org/10.1016/j.chroma.2021.462358.
[32] X. Guan, Z. Zhao, S. Cai, S. Wang, H. Lu, Analysis of volatile organic compounds using cryogen-free thermal modulation based comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry, Journal of Chromatography A 1587 (2019) 227–238. https://doi.org/10.1016/j.chroma.2018.12.025.
[33] L. Mikaliunaite, T.J. Trinklein, G.S. Ochoa, P.E. Sudol, D.S. Bell, R.E. Synovec, Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension, Journal of Chromatography Open 3 (2023) 100076. https://doi.org/10.1016/j.jcoa.2022.100076.
[34] C.E. Freye, L. Mu, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography, Journal of Chromatography A 1424 (2015) 127–133. https://doi.org/10.1016/j.chroma.2015.10.098.
[35] C.E. Freye, R.E. Synovec, High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry, Talanta 161 (2016) 675–680. https://doi.org/10.1016/j.talanta.2016.09.002.
[36] B.A. Parsons, D.K. Pinkerton, R.E. Synovec, Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, Journal of Chromatography A 1536 (2018) 16–26. https://doi.org/10.1016/j.chroma.2017.07.018.

Chapter 5. Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry

This chapter is reproduced from Lina Mikaliunaite, Robert E. Synovec, "Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry" *Talanta* 244 (2022) 123396.

5.1 Introduction

The study of the metabolome has been of immense interest in a variety of fields [1–5]. These samples are complex due to the high number of metabolites, eg., there are ~ 200,000 different metabolites in plants [6,7], and ~ 600 in yeast [8,9]. Metabolite sample complexity makes them well suited to gas chromatography (GC) analysis methods. However, challenges due to the high complexity and dimensionality of the datasets remain [10,11]. Fortunately, due to its high peak capacity and separation selectivity, comprehensive two-dimensional (2D) gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) [12–15] is well suited to address these challenges for metabolomics studies [16–20]. Additionally, chemometrics and various forms of data reduction can be applied to find statistically relevant analyte "features" within a sample or changes between different sample classes. In recent years, locating analyte peaks that indicate important changes between sample classes for GC×GC-TOFMS data has been performed using Fisher ratio (F-ratio) analysis [21,22]. The standard F-ratio [22] metric compares samples in one class to the other class(es), by taking the ratio of the between-class variance relative to the sum of the within-class variance. For metabolite studies, a high F-ratio often translates into the discovery of an important biological difference [21]. Software implementation of F-ratio analysis to GC×GC-TOFMS data has evolved from using a pixel-based F-ratio calculation, which is hampered by retention time variation on both the ^{1}D and ^{2}D separation dimensions [19,23], to

the current state-of-the-art in which a tile-based F-ratio algorithm is applied, which mitigates the retention time variation problem [24].

With the development of tile-based F-ratio analysis [24,25], and commercialization as ChromaTOF Tile, this data analysis tool readily facilitates a more thorough and enlightening examination of interesting datasets that were collected well in advance of this software development. Yeast metabolites, in particular, have been studied in the past two decades, with some research groups examining the relationship between oscillations in O_2 consumption, and periodicity in cellular yeast metabolite concentration [26–30]. Specifically, the yeast metabolomic cycle (YMC) was found to vary in length by different groups, and reports of ~5 hr [26] and 40 min [27] of length have been reported. In the current investigation, we revisit a previously collected YMC dataset for the diploid yeast strain CEN.PK that is based upon the ~5 hr cycle period, whereby samples were collected over two cycles (thus 10 hr) every 25 min while keeping the O_2 concentration oscillating [30], and (following sample preparation) analyzed by GC×GC-TOFMS. In this case, 24 different classes of samples were created (i.e., 24 time-point sequence over the 10 hr with a 25 min sampling interval), that were previously analyzed by Signal-ratio and PARAFAC software [31] to identify and quantify the discovered cycling metabolites [30]. While this cycling yeast dataset [30] was previously analyzed with some success, we now recognize that retention time variation (especially in the ^{2}D dimension) limited discovery of many cycling metabolites. Thus, reexamination of this cycling yeast dataset by tile-based F-ratio analysis is highly warranted to comprehensively discover all the cycling metabolites, while putting to the test the performance of the recently released ChromaTOF Tile software on a demanding dataset involving 24 sample classes.

Following discovery of the metabolites and generation of the F-ratio hitlist, a more challenging aspect of comprehensively analyzing cycling yeast metabolites is the development of a robust computational method that determines each cycling pattern, which is manifested in the signal sequence (i.e., concentration sequence) as a function of the sampling time-point. Previously this was performed by calculating the "depth-of-modulation," where the average of the two highest signals over the 24 sampling time-points was divided by the average of two lowest signals [30]. While this approach performed well for some metabolites, it did not account for the specific time spacing between crests and throughs of each data sequence, and random signal changes due to noise could cause two high concentration peaks and two low ones, resulting in a false positive. In the present study, we explore this previously obtained GC×GC-TOFMS YMC dataset using a computational model approach to facilitate a more accurate and comprehensive determination of what constitutes a "cycling" metabolite signal sequence for all of the hits in the tile-based F-ratio analysis hitlist.

Comparing experimental data to a computational model has been applied in a variety of arenas such as inorganic [32], environmental [33] and biological chemistry [34,35], whereby either the computational model is tested for accuracy using a "well controlled" experimental procedure, or in contrast, the computational model is used to decipher underlying relationship(s) in a given dataset. An example of the latter is fitting a sine wave to a data sequence to determine if temporal cycling is occurring [36]; if the data has been collected over a long period of time to generate several repeated oscillations, then fitting a sine wave often serves as an adequate model. In contrast, if the data sequences span a limited number of cycling periods, such as two cycles with the dataset in the current study, then such methods are not effective to determine if a given signal sequence can be said to be cycling. Furthermore, fitting a sine wave for the purpose of determining

if a dataset is cycling restricts the discovery to only the ideal cycling pattern of a sine wave, whereas with the dataset in the current study there is the potential for a variety of cycling patterns, so an untargeted method is needed.

To address this challenge and need, the primary goal of this study is to develop a computational method to comprehensively recognize and assign cycling data patterns in data sequences that may exhibit only two full cycles, and to accomplish this goal under conditions in which noise severely hampers the data interpretation. Broadly speaking, the method presented herein statistically compares the experimentally collected metabolite signal sequence data to randomly generated signal sequences to ultimately provide a confident, statistically based, identification of cycling metabolites. The first step is to apply tile-based F-ratio analysis via ChromaTOF Tile to a previously collected GC×GC-TOFMS YMC dataset [30]. Next, we take advantage of the experimental design that provided two full cycles for the yeast strain under investigation. We hypothesize that, independent of cycling phase, if a given metabolite exhibits a cycling pattern, then the pattern of the first half of the 24-point signal sequence should match reasonably well with the pattern of the second half from the signal sequence, i.e, the pattern should repeat. The similarity between the two half-sections is measured using a lack-of-fit (*LOF*) calculation [37], providing a quantitative metric. Specifically, the *LOF*, is calculated as follows,

$$LOF = 100\% \sqrt{\frac{\sum_i (y-x)_i^2}{\sum_i x_i^2}} \quad (1)$$

where x is the average peak area for each of the first 12 time-points (x is indexed from $i = 1$ to 12), and y is the average peak area for each of the last 12 time-points (y is indexed from $i + 12$ to 24) from a 24-point signal sequence. Thus, the peak area for time-point 1 is compared to 13, 2 to 14, and so on, to get a *LOF* for every hit of interest from the F-ratio hitlist. We also anticipate that the *LOF* will depend upon the relative standard deviation (*RSD*) of each 24-point metabolite signal

sequence, thus producing a *LOF* versus *RSD* distribution. Metabolites exhibiting a cycling pattern should exhibit a smaller *LOF*, while a larger *LOF* should be obtained due to a lack of a cycling pattern (or insufficient cycling pattern). A key challenge is to determine what magnitude of *LOF* could be obtained simply by random concentration changes (no cycling). To determine this *LOF* threshold, the method presented herein relies upon comparing the *LOF* versus *RSD* distribution obtained from all of the metabolites to a *LOF* versus *RSD* distribution obtained from randomly generated signal sequences. Metabolites that were found to exhibit cycling based upon having a sufficiently small, statistically significant *LOF* at a given *RSD* in relation to the *LOF* of randomly generated signal sequences were further examined to categorize their cycling pattern. Following categorization of the metabolite cycling patterns and phase, class assignments for all of the cycling metabolites were examined using principal component analysis (PCA) [38,39].

5.2 Experimental

In this study, we are analyzing chromatograms that were previously obtained using diploid yeast strain CEN.PK samples [26,30]. The chromatograms were collected using GC×GC-TOFMS employing the LECO Pegasus III system with a 4D upgrade (LECO, St. Joseph, MI, USA), which uses thermal modulation to transfer effluent from the ^{1}D column to the ^{2}D column. The GC×GC column set consisted of a ^{1}D column (20 m × 250 μm inside diameter × 0.5 μm film thickness) with an RTX-5MS stationary phase (Restek, Bellefonte, PA, USA), coupled by the modulator to a ^{2}D column (2 m × 180 μm inside diameter × 0.2 μm film thickness) with an RTX-200MS stationary phase (Restek, Bellefonte, PA, USA). The ^{1}D column was held at 60 °C for 0.25 min then ramped at 8 °C/min to 280 °C where it was held for 10 min, using a constant head flow rate of 1 ml/min. Analyte was modulated from the ^{1}D separation effluent onto the ^{2}D column using a modulation period of 1.5 s using a default hot pulse time of 0.4 s. The ^{2}D column was kept 10 °C

higher than the ^{1}D column, while the modulator was 40 °C higher than the ^{1}D column. The inlet and transfer line were kept constant at 280 °C and the TOFMS ion source was held at 200 °C. Yeast samples were collected over a 10 hr period every 25 min resulting in 24 time-points and following sample preparation that included derivatization by methoximation and trimethylsilylation [16]. Three injection replicates were collected using a 1 µl split-less injection at each time-point sample. The GC×GC-TOFMS data were collected for 32 min after a 5-min solvent delay (Fig. 5.1. A). Further experimental details on yeast metabolite sample preparation and separations was previously reported [29,30]. Data was imported into ChromaTOF Tile Software version 101 (LECO, St. Joseph, MI, USA) to perform tile-based F-ratio analysis, where the mass channel (*m/z*) that produced the highest F-ratio for each hit was used for hitlist ranking [40]. A more extensive summary of the parameters used in the implementation of ChromaTOF Tile software can be found in Table B.1. in the Supplementary Material. Artifact removal was applied to the F-ratio hitlist to make true positives move up toward the top of the hitlist, and then statistical thresholds were applied so as to only include metabolite hits that have a reasonable chance of exhibiting a cycling pattern in their signal sequence. For each metabolite hit passing these hitlist pruning steps, the replicates at each time-point were quantified using the ChromaTOF Tile software and then averaged. This quantification step resulted in a 24-point signal sequence (referred to as "signal sequence") using the top F-ratio *m/z* for each metabolite hit. Since not all of the metabolites discovered by tile-based F-ratio analysis were expected to cycle in a statistically meaningful way, further data analysis using these signal sequences to find cycling metabolites was performed using Matlab 2019b (Mathworks, Inc., Natick, MA, USA). All of the experimentally obtained metabolite signal sequences were normalized to have a mean of 1 to put on a level playing field to compare to the randomly generated signal sequences (i.e., simulated data). However, following

normalization, the metabolite signal sequences retained their native signal variability, which was quantified as the *RSD*.

All of the simulations of randomly generated signal sequences were performed in Matlab 2019b. Each sequence was 24 numbers in length, to correspond to the 24 data points for the normalized metabolite signal sequences and were created to follow a normal distribution. The randomly generated signal sequences were created over a range of *RSD*, from 10% to 100% every 5% with a mean of 1. At each of the *RSD*, 10000 sequences were created, for a total of 190,000 randomly generated signal sequences. Confidence level curve fitting was performed in Matlab 2019b using Curve Fitting Toolbox (Mathworks, Inc., Natick, MA, USA) and PCA plots were made using PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA).

The metabolite cycling patterns will be shown to fit into four category patterns: 1 cycle/5 hr period metabolites, 2 cycles/5 hr period metabolites, "spiky" metabolites, or multimodal metabolites. Although the use of a sine wave model was not effective to obtain a comprehensive list of all cycling metabolites, categorization of the 1 cycle/5 hr period metabolites and 2 cycles/5 hr period metabolites from all of the cycling metabolites was facilitated by comparing those metabolite signal sequences to a sine wave model coupled with the same *RSD* as the given metabolite. After confidently assigning 1 cycle/5 hr period metabolites and 2 cycles/5 hr period metabolites, spiky metabolites were found by manually inspecting the remaining metabolite patterns, specifically seeking to find the metabolites that exhibited two high signals separated exactly 12 time-point intervals apart (i.e., 1 cycle). Additionally, phase assignments for the 1 cycle/5 hr period metabolites, 2 cycles/5 hr period metabolites and some of spiky metabolites were obtained by shifting the sine wave model over relative to the normalized metabolite signal sequence to find the lowest *LOF* which defined the best phase assignment. Finally, the multimodal

metabolite patterns were observed to have less obvious sine-like patterns that included cycles with a period less than 2.5 hr.

5.3 Results and Discussion

Figure 5.1. A shows a *m/z* 73 chromatogram of a yeast sample, while Fig. 5.1 B provides a zoom in region of the chromatogram so the separation detail can be better visualized. Tile-based F-ratio analysis was performed on 72 chromatograms, which were split into the 24 time-point classes, with 3 chromatograms per class (3 replicate injections). The ChromaTOF Tile software found 2818 hits, out of which 2207 hits had a p-value < 0.05. Following removal of artifact peaks due to the derivatization process and redundant hit removal, 672 hits remained. Next, using the top F-ratio *m/z* for each of the 672 hits, the replicates at each time-point were averaged to reduce the data density from 72 points to the final 24-point signal sequences for each hit, with an example of the averaging process for the myo-inositol peak area signals provided in Fig. B.1.

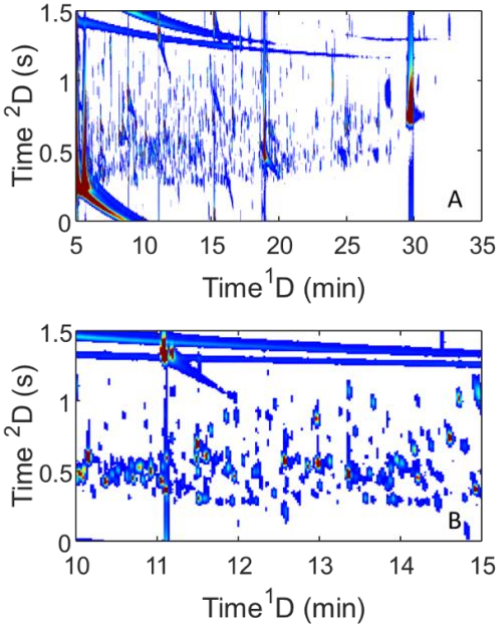


Figure 5.1. (A) GC×GC-TOFMS chromatogram of yeast metabolite sample using m/z 73, where the chromatogram on ^{1}D starts at 5 min, to account for the solvent delay. (B) A zoomed-in portion of the chromatogram to facilitate visualizing the separation.

Examples of metabolite cycling patterns in the context of the changes in dissolved O_2 content over the 2 cycles (5 hr per cycle period) of sample collection are provided in Fig. 5.2. A variety of metabolite cycling patterns were observed, such as myo-inositol (Fig. 5.2. A) which exhibits a sinusoidal pattern in its 24 time-point signal sequence. In contrast, methyl citrate (Fig. 5.2. B) exhibits what we refer to as a "spiky" cycling pattern, where there are two high intensity signals separated by a 12 time-point interval (exactly 1 cycling period), suggesting the metabolite concentration is high only for a short period of time within a cycling period. As another example,

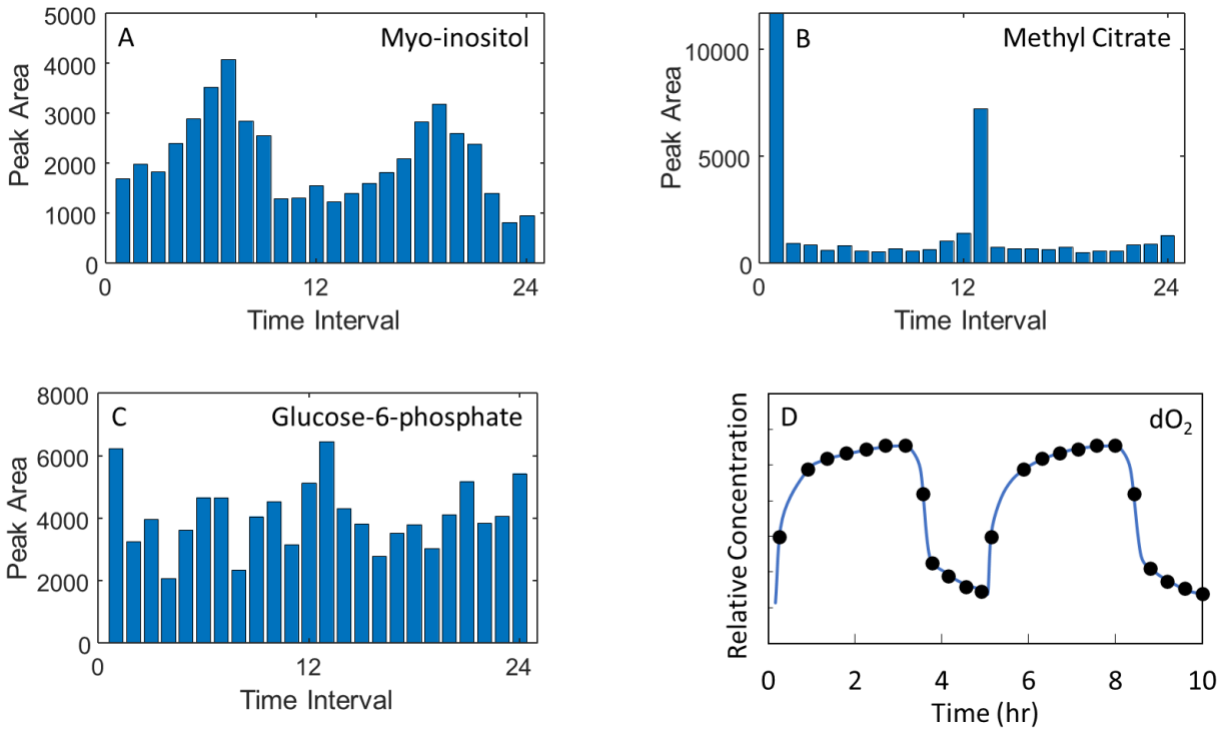


Figure 5.2. Examples of the metabolite cycling patterns in the context of the changes in dissolved O2 level over the 2 cycles (5 hr per cycle period) for the diploid yeast strain CEN.PK. (A) Myo-inositol exhibits a 1 cycle/5 hr cycle sinusoidal pattern in its 24 time-point signal sequence. (B) Methyl citrate exhibits a "spiky" cycling pattern, where there are two high intensity signals separated by a 12 time-point interval (exactly 1 cycling period). (C) Glucose-6-phosphate exhibits multimodal cycling, where it will be shown there is a correlation between the first half (time-points 1 – 12) and the second half (time-points 13-24) of the 24 time-point signal sequence. (D) Relative O2 concentration oscillation pattern, where the 24 black dots indicate when the yeast samples were collected (25 min apart).

glucose 6-phosphate in Fig. 5.2 C might at first glance seem without a clear cycling pattern, but will be demonstrated to exhibit multimodal cycling, where there is a correlation between the first

half (time-points 1 – 12) and the second half (time-points 13-24) of the 24 time-point signal sequence, but within one ultradian YMC [26]; thus, glucose 6-phosphate expresses more than 1 cycle during each 5 hr period. All of these cycling patterns are caused by changes in dissolved O_2 content that had 2 cycles within 10 hr of sample collection as indicated provided in Fig. 5.2. D, where black dots indicate at what time within the O_2 cycle the yeast samples were collected.

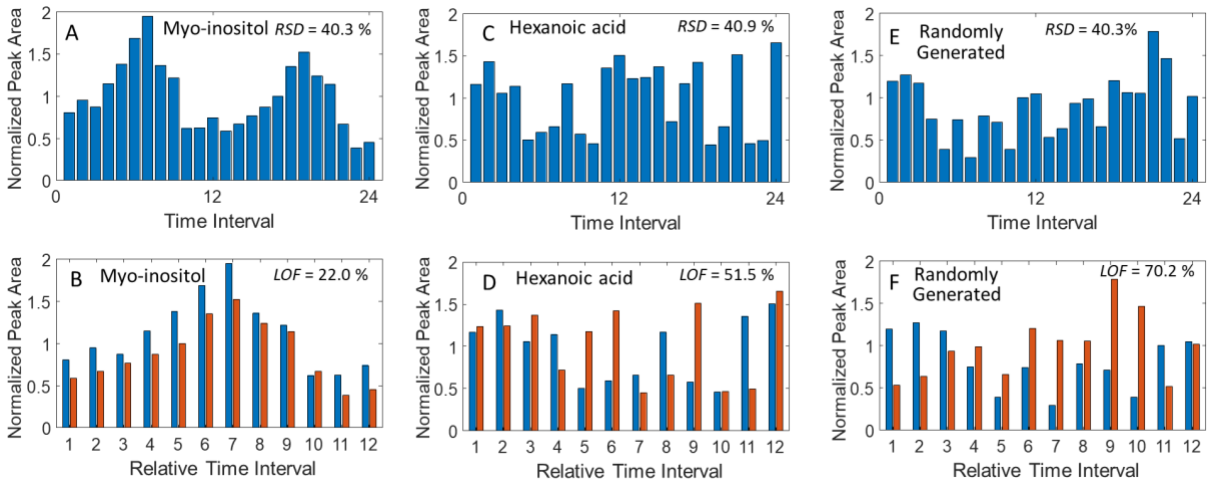


Figure 5.3. The process for calculating the RSD and LOF metrics to find cycling metabolites is illustrated by examples. (A) Myo-inositol 24 time-point signal sequence with an RSD of 40.3%. (B) Determination of LOF for myo-inositol where the first 12 time-points (1 – 12) and the last 12 time-points (13 – 24) from (A) are matched up 1 to 13, 2 to 14, etc., and then Eq. (1) is applied, resulting in a LOF of 22.0%. (C) Hexanoic acid 24 time-point signal sequence with an RSD of 40.9%. (D) Determination of the LOF for hexanoic acid resulting in a LOF of 51.5%. (E) Randomly generated signal sequence with an RSD of 40.3%. (F) Determination of the LOF for the randomly generated signal sequence in (E) resulting in a LOF of 70.2%.

While ChromaTOF Tile found 672 metabolites that had significant changes between the 24 time-point classes ($p < 0.05$), these hits needed to be rigorously evaluated for differences indicative of cycling relative to random concentration changes. This process of determining the *LOF* via Eq. (1) is presented in Figs. 5.3. A,B for myo-inositol, Figs. 5.3. C,D for hexanoic acid, and Figs. 5.3. E,F for a randomly generated signal sequence. We expect a cycling metabolite would have a small *LOF*, whereas a non-cycling metabolite that exhibited essentially a random signal pattern and a randomly generated signal sequence, would both exhibit a large *LOF*. For example, in Figs. 5.3. A,B for myo-inositol produces a *LOF* of 22.0%, whereas another hit, identified as

hexanoic acid, does not have a clear cycling pattern (Figs. 5.3. C,D), producing a *LOF* more than twice as large at 51.5%. Since the *LOF* calculation can be performed on all 672 of the hits, and visually it is highly suggestive that a smaller *LOF* might correspond to biologically relevant cycling in the metabolite concentration, a more accurate and robust determination of the *LOF* threshold below which a hit would be "quantitatively" defined as cycling still has to be determined. To do so, randomly generated signal sequences were created (example in Fig. 5.3 E), which mimic non-cycling real data. The *LOF* between the first cycle (time-points 1 – 12) and second cycle (time-points 13 – 24) are calculated the same way for the randomly generated signal sequences as for the metabolite hits, as demonstrated in Fig. 5.3. F with a *LOF* of 70.2%.

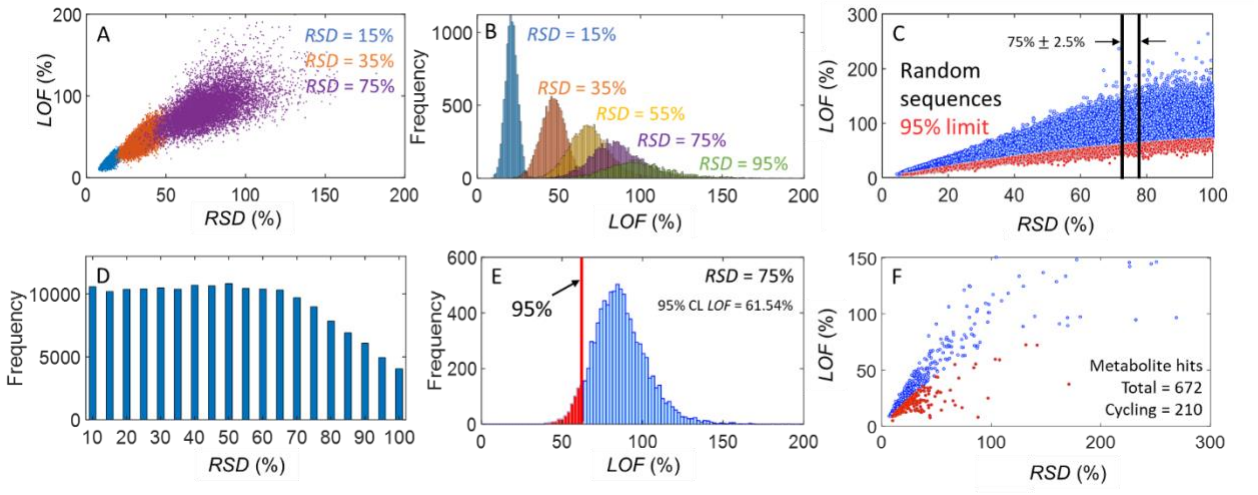


Figure 5.4. Steps in the computational method to identify cycling metabolites. (A) Overlay of the observed 2D distributions of LOF versus RSD, for three simulated RSD levels: 15%, 35% and 75%. There are 10000 simulations at each simulated RSD level. (B) Frequency of the LOF obtained for the RSD simulated at 15%, 35%, 55%, 75% and 95%. (C) Overlay of all 2D distributions of LOF versus RSD for the randomly generated signal sequences. There were 10000 randomly generated signal sequences created at each RSD from 10% to 100% spaced every 5%, resulting in 19 simulated RSD conditions or a total of 190,000 sequences. The "slicing method" is illustrated at an RSD of 75% ±2.5%, which contained 8996 LOF versus RSD values. The LOF threshold at the 95% confidence level ($p < 0.05$) is indicated by the transition from blue to red dots. (D) Frequency of LOF for each centered RSD level as illustrated for RSD of 75% in (C). (E) Illustration of the 95% confidence level for the RSD centered at 75% in (C). (F) Application of the 95% confidence level generated in (C) to the metabolite signal sequences. The 672 metabolite hits discovered by F-ratio analysis divide into two groups: (red dots) 210 metabolites that exhibit non-random (cycling) behavior, and (blue dots) 462 metabolites that are not cycling.

In order to provide an accurate determination of the *LOF* cycling pattern threshold, randomly generated signal sequences span a range of *RSD* to match the range of the experimental

GC×GC-TOFMS data, normally distributed to have a mean of 1, as the mean of the data has no impact over *LOF* (Fig. B.2.). To adequately account for the full range of possible random sequences, 10000 of the randomly generated signal sequences were created at each *RSD* from 10% to 100% spaced every 5%, resulting in 19 simulated *RSD* conditions or a total of 190,000 sequences. Although the randomly generated signal sequences were created at 5% *RSD* intervals, the resulting "observed" *LOF* and *RSD* for each simulated *RSD* produce *2D distributions* of *LOF* versus *RSD*, with the results at 3 of the 19 *RSD* simulated conditions presented in Fig. 5.4. A for a simulated *RSD* of 15%, 35% and 75%. Another way to view the simulated data is to plot the frequency of the *LOF* obtained for a given *RSD*, as presented in Fig. 5.4. B for the *RSD* simulated at 15%, 35%, 55%, 75% and 95%; the distributions are Gaussian, consistent with the simulated conditions. The distributions shown in Figs. 5.4. A,B indicate that standard deviation of the *LOF* increases as the observed *RSD* increases, both increasing as the nominal *RSD* input to the simulation increases. Thus, due to the shape of these *2D distributions* of *LOF* versus simulated *RSD* (per Fig. 5.4. A), in order to obtain an accurate determination of the localized *LOF* for a either the observed simulated or experimentally obtained *RSD*, all of the 19 *2D distribution*s of *LOF* versus *RSD* for the randomly generated signal sequences were combined as presented in Fig. 5.4. C. As a quality control measure to ensure that 5% *RSD* intervals provided sufficient overlap along the *RSD* axis, the frequency of *LOF* versus *RSD* in 5% *RSD* intervals were counted, centered at the same intervals as the original simulated *RSD* by ±2.5%. This "slicing method" exercise to determine the localized *LOF* frequency per *RSD* is illustrated in Fig. 5.4 C at the *RSD* of 75%, which contained 8996 *LOF* versus *RSD* values. The resulting plot of frequency of *LOF* versus *RSD* are provided in Fig. 5.4. D, where we see the frequency is relatively constant at ~ 10000 up to *RSD* of ~ 70% where the frequency begins to smoothly taper to a lower frequency of *LOF*. The constant

and/or smooth distribution in Fig. 5.4. D ensures that a suitable "statistically-based" *LOF* metric can be determined along the *RSD* axis for the purpose of confidently determining whether or not a given metabolite hit is exhibiting a signal sequence that contains information indicative of cycling. For example, as illustrated in Fig. 5.4. E for the frequency of *LOF* versus *RSD* results centered at an *RSD* of 75% (centered on a 5% *RSD* interval), the 95% confidence level ($p < 0.05$) requires the $LOF \geq 61.54\%$ to indicate the signal sequence pattern originated from randomly generated data. Thus, conversely, a $LOF < 61.54\%$ at an *RSD* of 75% indicates the signal sequence contains sufficient evidence of non-random information, i.e., cycling at some level. Application of this exercise to determine the 95% confidence level at all *RSD* was then performed with the results indicated in Fig. 5.4. C by the red dots (band of lower *LOF*). It is important to note that the *LOF* versus *RSD* does not exhibit a linear relationship (more details in Fig. B.3.) which is why the "slicing method" presented in Fig. 5.4. C had to be implemented to determine an accurate *LOF* threshold. While this *LOF* versus *RSD* statistical threshold determination was nominally applied from a simulated *RSD* range of 10% to 100%, the resulting 95% confidence interval for *LOF* were readily fit to an equation and extrapolated to slightly higher *RSD* (Fig. B.3.) for application to a few of the experimentally obtained signal sequences. Next, for the 672 metabolites that had significant changes between the 24 time-point classes (i.e., $p < 0.05$), each of them had their *LOF* and *RSD* determined as shown in Fig. 5.3. for myo-inositol and hexanoic acid. Of the 672 metabolites, 210 possessed a $LOF < LOF$ threshold ($p < 0.05$) required to indicate cycling behavior. These results are presented in Fig. 4F, which implements the *LOF* threshold illustrated in Fig. 5.4. C and Fig. B.3.

After determining the 210 metabolites that possessed a *LOF* for their *RSD* indicative of a statistically significant level of cycling behavior, the next step was to determine the cycling

category and, if applicable, the phase relative to the O_2 cycling pattern (Fig. 5.2. D). Four metabolite cycling pattern categories were observed: 1 cycle/5 hr period, 2 cycles/5 hr period, "spiky", or multimodal. The *LOF* and *RSD* metrics were used again but applied in a different manner to find metabolites that had either 1 or 2 cycles. Starting with the 210 metabolites, to test for a 1 cycle/5 hr period pattern, a sine wave model of 24 points was constructed with the same standard deviation (*SD*) and average peak area signal as the given metabolite data, as presented in Fig. 5.5. A for myo-inositol. Next, the *LOF* between the sine wave model and metabolite data was repeatedly calculated via Eq. (1) by moving the sine wave one time-point over (Fig. 5.5. B) where the crest moving by one time-point to the left accounts for a 30° phase change (i.e., 360° in 1 cycle is defined by 12 time-points, so 360 ° /12 = 30 °), and 0° was assigned according to previously published work to coincide with the O_2 cycle [30].

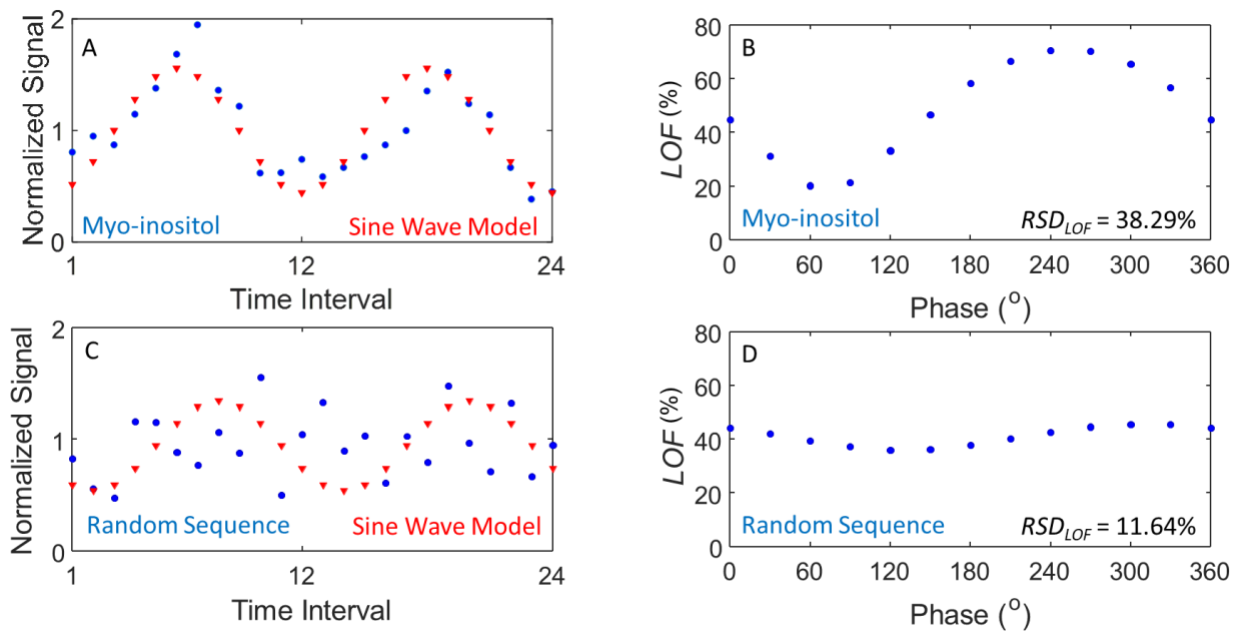


Figure 5.5. Metabolite cycling phase determination process. (A) The 1 cycle/5 hr period sine wave model (red triangles) of 24 time-points (2 cycles over 10 hr) was constructed with the same SD as the signal sequence for myo-inositol (blue dots). (B) The LOF between the sine wave model and myo-inositol signal sequence was repeatedly calculated via Eq. (1) by moving the sine wave one time-point over, i.e., 30° phase change per time-point, with an RSD_{LOF} of 38.29% obtained. (C) A randomly generated signal sequence (blue dots) is shown in relation to a sine wave model (red triangles) with the same SD. (D) The LOF between the sine wave model and randomly generated signal sequence is shown, with an RSD_{LOF} of 11.64% obtained.

There are 12 different possible phases, 0° to 330°, where the lowest *LOF* of the twelve indicates the phase that best matches a given metabolite. The applied metric for the cycling phase determination was the *RSD* of the *LOF* plot, or simply the RSD_{LOF}. For myo-inositol the RSD_{LOF} was 38.29%. While the phase determination for a 1 cycle/5 hr period metabolite such as myo-inositol in Figs. 5.5. A,B is easily determined, there is more ambiguity for many of the other metabolites that nonetheless passed the initial test of exhibiting some level of cycling (non-random temporal behavior). Thus, in an analogous fashion, the previously applied randomly generated signal sequences were employed to facilitate a statistically based determination of the cycling phase for the 1 cycle/5 hr period and 2 cycles/5 hr period metabolites. An example of the determination of the cycling phase (more specifically the lack thereof) for a randomly generated signal sequence is presented in Figs. 5.5. C,D. Note that, as with the experimental data, the sine wave model has the same *SD* as the randomly generated signal sequence. Due to the randomness of the sequence in Fig. 5.5. C, the *LOF* plot in Fig. 5.5. D calculated by iteratively shifting the sine wave model relative to the randomly generated signal sequence by one time-point intervals (30° phase change) exhibits a very small variation in *LOF*, quantified by a RSD_{LOF} of 11.64%, which is lower than the RSD_{LOF} of 38.29% obtained for myo-inositol. The notion is that a sufficiently high RSD_{LOF} threshold must be exceeded in order to confidently assign a cycling phase for an experimentally obtained signal sequence relative to that of the randomly generated signal sequences. We explore the determination of this threshold next.

The RSD_{LOF} were calculated for all metabolites and for the 190,000 randomly generated signal sequences, following the approach illustrated in Fig. 5.5. Note that while the *LOF* as a function of *RSD* was non-linear in the determination of cycling behavior per Fig. 4 that produced shifting *LOF* distributions as a function of *RSD*, the determination of the phase using the RSD_{LOF}

metric as in Fig. 5.5 is naturally normalized so a single distribution is obtained, one each for the randomly generated signal sequences (Fig. 5.6. A) and the metabolite signal sequences (Fig. 5.6. B). Indeed, even though the *SD* was matched for the sake of visualization in Fig. 5.5, the calculation of the RSD_{LOF} metric is independent of the initial *SD* for the randomly generated signal sequences and the metabolite signal sequences. The RSD_{LOF} distribution for the randomly generated signal sequences is concentrated to relatively small RSD_{LOF} (Fig. 5.6. A), and a 95% confidence level ($p < 0.05$) for RSD_{LOF} of 19.16% was determined. In contrast, the RSD_{LOF} distribution for the metabolite signal sequences (Fig. 5.6. B) is shifted to much higher RSD_{LOF} with 104 exceeding the RSD_{LOF} of 19.16% threshold out of the 210 total metabolites that were found to exhibit some level of cycling (Fig. 5.4. F). Thus, 104 out of 210 metabolites were categorized as 1 cycle/5 hr period metabolites with their phase also confidently assigned.

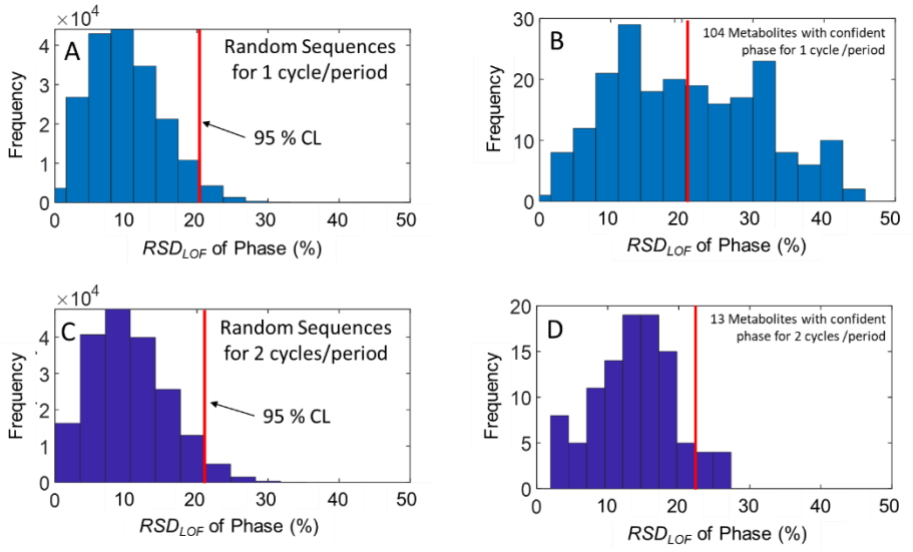


Figure 5.6. Statistically based cycling and phase assignment for the 1 cycle/5 hr period and 2 cycle/5 hr period metabolites. (A) Distribution of RSD_{LOF} for the 190000 randomly generated signal sequences using the 1 cycle/5 hr period sine wave model (eg., Fig. 5.5. D). The 95% confidence level ($p < 0.05$) at an RSD_{LOF} of 19.16% was determined (red line). (B) Distribution of RSD_{LOF} for the 210 metabolite signal sequences (eg., Fig. 5.5. B) in which 104 exceed the RSD_{LOF} of 19.16% threshold. (C) Distribution of RSD_{LOF} for the 190000 randomly generated signal sequences using a 2 cycle/5 hr period sine wave model. The 95% confidence level ($p < 0.05$) at an RSD_{LOF} of 20.19% was determined (red line). (D) Distribution of RSD_{LOF} for the 210 metabolite signal sequences in which 13 exceed the RSD_{LOF} of 20.19% threshold.

A similar approach was applied to the remaining 106 out of 210 metabolites to assess which could be categorized as 2 cycles/5 hr period metabolites with an accompanying phase determination. For this step, 6 different phase assignments were possible, every 60^{o} from 0 to 330^{o}. The distribution for the randomly generated signal sequences treated as if they were 2 cycles/5 hr period metabolites (Fig. 5.6. C) did not appear substantially different then the distribution for the remaining 106 out of 210 metabolite signal sequences (Fig. 5.6. D). However, at 95% confidence level for RSD_{LOF} of 20.19%, 13 metabolites were categorized as having a 2 cycles/5 hr period and phase assignment. Next, the remaining 93 out of 210 metabolites were manually inspected to find those exhibiting a spiky pattern, as they have an easily identifiable pattern of two high intensity peaks exactly 12 time-point intervals apart. After the inspection, 12 metabolites were identified as having a spiky pattern and their phase assigned by comparison to a sine wave model; however, a few of the metabolites with a spiky pattern exhibited anomalous patterns so their phase was not assignable. The remaining 79 metabolites were deemed to exhibit a multimodal cycling pattern, and a phase could not be confidently assigned to them. These 79 metabolites exhibited a signal sequence pattern that possessed a statistically sufficient non-random behavior to be categorized as cycling per the method illustrated in Fig. 5.4., however their cycling behavior is not evident by visual inspection.

Table 5.1. The 55 identified metabolites were arranged according to their cycling category: 1 cycle/5 hr period (1c), 2 cycles/5 hr period (2c), spiky (sp), or multimodal (mm). A star (*) near the analytes indicates that were previously found [30]. For every cycling category except multimodal, the metabolites were ordered according to phase in degrees. An accompanying heatmap is provided in Fig. 5.7.

Phase	**Name**		**Phase**	**Name**
1c, 0	aspartic acid *		1c, 180	L-lysine
1c, 0	L-spartate		1c, 180	phenylethanolamine
1c, 0	propanetriol, 2-methyl-		1c, 210	lysine *
1c, 0	L-aspartic acid		1c, 210	DL-lactide
1c, 30	2-hydroxyisovaleric acid		1c, 210	DL-ornithine *

1c, 30	L-threonic acid *		1c, 240	citric acid
1c, 30	L-norvaline *		1c, 270	6-phospho-d-gluconate *
1c, 30	phenylacetic acid		2c, 210	D-ribulose 5-phosphate *
1c, 30	L-asparagine		sp, 120	fructose 1,6 bisphosphate
1c, 30	2-OH-isovaleric acid *		sp, 120	glycerol 3-phosphate *
1c, 60	butane *		sp, 120	glycerone phosphate
1c, 60	myo-inositol *		sp, 270	methyl citrate*
1c, 60	L-isoleucine *		sp, m	glycerol *
1c, 60	L-threonine *		mm	D-glucose 1-phosphate *
1c, 60	L-tyrosine		mm	propionic acid
1c, 90	L-valine *		mm	mannitol *
1c, 120	glycine *		mm	ribitol *
1c, 120	alanine *		mm	stearic acid
1c, 120	D-mannose		mm	ethanolamine
1c, 150	I-serine *		mm	glycine
1c, 150	2-ketoglutaric acid		mm	triethylene glycol bis(trimethylsilyl) ether
1c, 150	L-leucine *		mm	L-glutamic acid
1c, 150	L-proline *		mm	D-ribose 5-phosphate *
1c, 180	L-homoserine *		mm	propionylglycine
1c, 180	pyruvate *		mm	NAD+ *
1c, 180	L-glutamine		mm	cyclohexane, 1-methyl-2-propyl-
1c, 180	cystathionine tetra- *		mm	glucose-6-phosphate *
1c, 180	malonic acid			

For the 210 cycling metabolites discovered, 55 were tentatively identified (Table 5.1) based upon possessing both a forward and reverse mass spectrum match value above 800, although matching them with retention indexes was not performed, leaving 155 unknown cycling metabolites. While a majority of the cycling metabolites were not identified, likely since many derivatized metabolites are thought to not be in the mass spectrum library, their identification was not a goal of this study [30,41]. The 55 identified metabolites were arranged according to their cycling category in Table 5.1.: 1 cycle/5 hr period (1c), 2 cycles/5 hr period (2c), spiky (sp), or

multimodal (mm). All metabolite cycling categories were ordered according to phase except multimodal. A heatmap for the identified metabolites from Table 5.1. is provided in Fig. 5.7., where the arrangement by cycling category and phase. Inspection of the heatmap supports the notion that the methodology developed to identify cycling metabolites was effective. Similarly, the results for the 155 unknown metabolites are provided in Table B.2. and Fig. B.4., which reinforces the findings for the identified metabolites. In our previous study [30], 44 identified and 41 unknown for a total of 85 cycling metabolites were discovered, which means 125 cycling metabolites of the total 210 cycling metabolites are newly found using ChromaTOF Tile and then further categorized by the method presented herein.

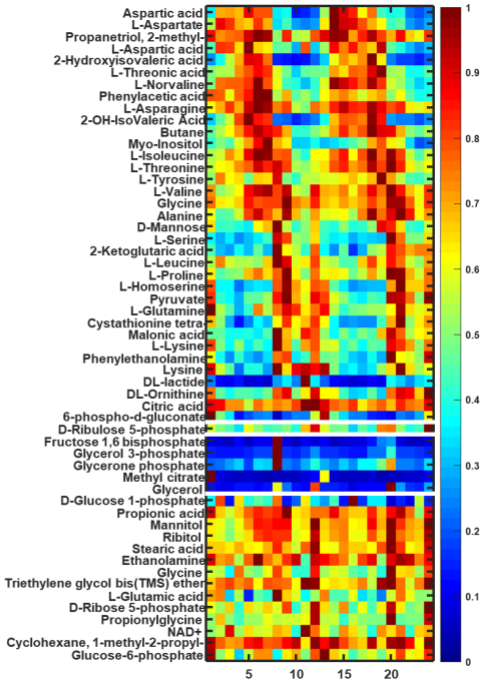


Figure 5.7. Heatmap for the 55 identified metabolites (per Table 5.1.), arranged by cycling category: first 35 metabolites are 1 cycle/5 hr period (1c), separated by white line from D-ribulose 5-phosphate which is the only identified metabolite of the 2 cycles/5 hr period (2c) type, separated with a white line from 5 spiky (sp) metabolites that are separated by another white line from the remaining 14 multimodal (mm) metabolites. Every category except multimodal is ordered according to phase. The red color indicates a high concentration and blue identifies a low concentration. Results for the 155 unknown metabolites are provided in Table B.2. and Fig. B.4.

To further visualize the categorization of the 210 cycling metabolites, their 24 time-point signal sequences were subjected to PCA with the scores plot provided in Fig. 5.8. A, where PC1

captures 51.1% of the variance and PC2 captures 18.2% of the variance. The four cycling metabolite categories (1 cycle/5 hr period, 2 cycles/5 hr period, spiky, or multimodal) cluster into readily interpretable groups. Metabolites with 1 cycle/5 hr period signal sequences are heavily reliant up PC2, with this observation reinforced by the shape of the loadings plot for PC2 (Fig. 5.8. B) along with the high confidence of the phase determination method employed in Figs 6A,B. Meanwhile, the metabolites with 2 cycles/5 hr period signal sequences are much less reliant upon PC2, which is consistent with their phase assignment being very near the statistical limit of the phase determination method employed in Fig. 5.6. C,D. Also, metabolites with spiky patterns fall to the left side of PC1, while essentially being invariant with PC2. Finally, the metabolites with multimodal pattern are most similar to the metabolites with 2 cycles/5 hr period.

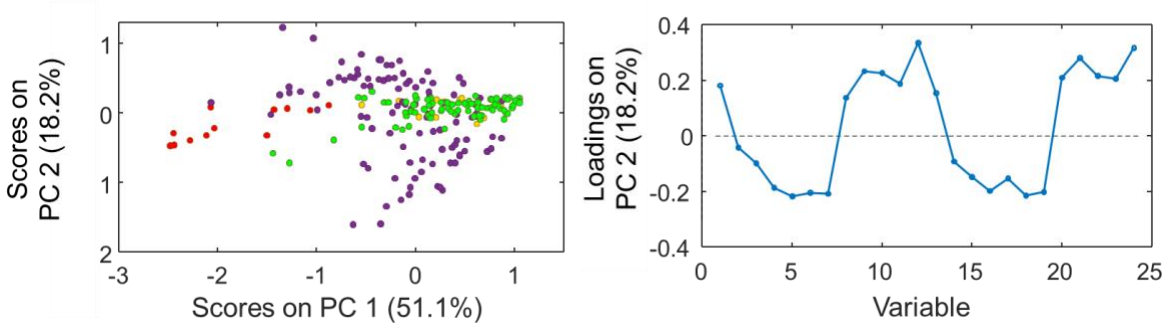


Figure 5.8. (A) PCA scores plot for all 210 cycling metabolites labeled by cycling category: 1 cycle/5 hr period (purple), 2 cycles/5 hr period (yellow), spiky (red), or multimodal (green). (B) PC2 loadings, showing a repeating cycling pattern between first 12 time-points and last 12 time-points.

5.4 Conclusion

A computational method for the comprehensive and untargeted determination of cycling yeast metabolites using GC×GC-TOFMS has been presented. The yeast metabolomic cycle for the diploid yeast strain CEN.PK based upon a 5 hr cycle period was thoroughly examined to comprehensively determine all of the discoverable metabolites that exhibited cycling. A key challenge for this study was that the samples were collected over only two cycles. Due to the

limited number of cycles expressed in the system, a computational method needed to be devised to determine with statistical significance whether or not a given metabolite was exhibiting a temporal signal pattern that constituted cycling in the context of the 5 hr cycle period. More broadly this challenge would occur as a consequence of not being able to collect samples over more cycle periods to time due to experimental and/or resource constraints. The computational method we developed compares the experimentally obtained metabolite signal sequences to randomly generated signal sequences coupled with a statistically based confidence level *LOF* metrics to determine whether or not a given metabolite expresses cycling, and if so, what is the phase of the cycling. Initially the GC×GC-TOFMS dataset was analyzed using ChromaTOF Tile, which performed tile-based F-ratio analysis. Since there were 24 time-point intervals, this constituted 24 sample classes in the F-ratio calculation which produced 672 metabolite hits. Next, application of the computational method presented herein using randomly generated signal sequences determined that there were 210 metabolites exhibiting signal sequence patterns containing cycling information: 55 identified metabolites and 155 unknown metabolites. The 210 cycling metabolites discovered is substantially higher than the 85 cycling metabolites that were found when the dataset was analyzed using less developed software tools [30]. Furthermore, the 210 cycling metabolites were readily categorized into four groups, and where applicable, a phase determined relative to the O_2 level in the yeast growth conditions: 1 cycle/5 hr period (35 identified and 71 unknown), 2 cycles/5 hr period (1 identified and 12 unknown), spiky (5 identified and 7 unknown), or multimodal (14 identified and 65 unknown). Moreover, the computational method of using random number sequences to find cycling metabolites presented herein is not limited to GC×GC-TOFMS data, or even metabolomics. Indeed, application of this method for confident identification of cycling analytes is broad in scope and is well suited for noisy data (as

demonstrated herein), and the only requirement is that at least 2 cycles worth of data have been collected. Thus, this computational method is well suited for the study of low concentration analytes, or when less sensitive detection methods are used.

5.5 References

[1] J.J. Kellogg, T.N. Graf, M.F. Paine, J.S. McCune, O.M. Kvalheim, N.H. Oberlies, N.B. Cech, Comparison of metabolomics approaches for evaluating the variability of complex botanical preparations: green tea (Camellia sinensis) as a case study, J. Nat. Prod. 80 (2017) 1457–1466. https://doi.org/10.1021/acs.jnatprod.6b01156.

[2] Edna.P. Nyangale, Donald.S. Mottram, Glenn.R. Gibson, Gut microbial activity, implications for health and disease: the potential role of metabolite analysis, J. Proteome Res. 11 (2012) 5573–5585. https://doi.org/10.1021/pr300637d.

[3] E.J. Want, A. Nordström, H. Morita, G. Siuzdak, From exogenous to endogenous: The inevitable imprint of mass spectrometry in metabolomics, J. Proteome Res. 6 (2007) 459–468. https://doi.org/10.1021/pr060505+.

[4] S. Moco, B. Schneider, J. Vervoort, Plant micrometabolomics: the analysis of endogenous metabolites present in a plant cell or tissue, J. Proteome Res. 8 (2009) 1694–1703. https://doi.org/10.1021/pr800973r.

[5] A. Bhide, A. Ganguly, T. Parupudi, M. Ramasamy, S. Muthukumar, S. Prasad, Next-generation continuous metabolite sensing toward emerging sensor needs, ACS Omega 6 (2021) 6031–6040. https://doi.org/10.1021/acsomega.0c06209.

[6] E. Pichersky, D.R. Gang, Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective, Trends Plant Sci. 5 (2000) 439–445. https://doi.org/10.1016/S1360-1385(00)01741-6.

[7] R.J. Bino, R.D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B.J. Nikolau, P. Mendes, U. Roessner-Tunali, M.H. Beale, R.N. Trethewey, B.M. Lange, E.S. Wurtele, L.W. Sumner, Potential of metabolomics as a functional genomics tool, Trends Plant Sci. 9 (2004) 418–425. https://doi.org/10.1016/j.tplants.2004.07.004.

[8] J. Förster, I. Famili, P. Fu, B.Ø. Palsson, J. Nielsen, Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network, Genome Res. 13 (2003) 244–253. https://doi.org/10.1101/gr.234503.

[9] W.B. Dunn, N.J.C. Bailey, H.E. Johnson, Measuring the metabolome: current analytical technologies, Analyst 130 (2005) 606–625. https://doi.org/10.1039/B418288J.

[10] P. Jonsson, A.I. Johansson, J. Gullberg, J. Trygg, J. A, B. Grung, S. Marklund, M. Sjöström, H. Antti, T. Moritz, High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses, Anal. Chem. 77 (2005) 5635–5642. https://doi.org/10.1021/ac050601e.

[11] K. Hollywood, D.R. Brison, R. Goodacre, Metabolomics: current technologies and future trends, Proteomics 6 (2006) 4716–4723. https://doi.org/10.1002/pmic.200600106.

[12] S.E. Prebihalo, K.L. Berrier, C.E. Freye, H.D. Bahaghighat, N.R. Moore, D.K. Pinkerton, R.E. Synovec, Multidimensional gas chromatography: advances in instrumentation, chemometrics, and applications, Anal. Chem. 90 (2018) 505–532. https://doi.org/10.1021/acs.analchem.7b04226.

[13] M.S.S. Amaral, Y. Nolvachai, P.J. Marriott, Comprehensive two-dimensional gas chromatography advances in technology and applications: biennial update, Anal. Chem. 92 (2020) 85–104. https://doi.org/10.1021/acs.analchem.9b05412.
[14] C. Cordero, H.-G. Schmarr, S.E. Reichenbach, C. Bicchi, Current developments in analyzing food volatiles by multidimensional gas chromatographic techniques, J. Agric. Food Chem. 66 (2018) 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997.
[15] P.J. Marriott, S.-T. Chin, Y. Nolvachai, Techniques and application in comprehensive multidimensional gas chromatography – mass spectrometry, J. Chromatogr. A 1636 (2021) 461788. https://doi.org/10.1016/j.chroma.2020.461788.
[16] R.E. Mohler, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells, Anal. Chem. 78 (2006) 2700–2709. https://doi.org/10.1021/ac052106o.
[17] M.F. Almstetter, I.J. Appel, M.A. Gruber, C. Lottaz, B. Timischl, R. Spang, K. Dettmer, P.J. Oefner, Integrative normalization and comparative analysis for metabolic fingerprinting by comprehensive two-dimensional gas chromatography−time-of-flight mass spectrometry, Anal. Chem. 81 (2009) 5731–5739. https://doi.org/10.1021/ac900528b.
[18] M. Menéndez-Carreño, H. Steenbergen, H.-G. Janssen, Development and validation of a comprehensive two-dimensional gas chromatography-mass spectrometry method for the analysis of phytosterol oxidation products in human plasma, Anal. Bioanal. Chem. 402 (2012) 2023–2032. https://doi.org/10.1007/s00216-011-5432-2.
[19] H.D. Bean, J.E. Hill, J.-M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography–mass spectrometry data, J. Chromatogr. A 1394 (2015) 111–117. https://doi.org/10.1016/j.chroma.2015.03.001.
[20] F.A. Franchina, D. Zanella, T. Dejong, J.-F. Focant, Impact of the adsorbent material on volatile metabolites during in vitro and in vivo bio-sampling, Talanta 222 (2021) 121569. https://doi.org/10.1016/j.talanta.2020.121569.
[21] S.E. Prebihalo, G.S. Ochoa, K.L. Berrier, K.J. Skogerboe, K.L. Cameron, J.R. Trump, S.J. Svoboda, J.K. Wickiser, R.E. Synovec, Control-normalized fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data for enhanced biomarker discovery in a metabolomic study of orthopedic knee-ligament injury, Anal. Chem. 92 (2020) 15526–15533. https://doi.org/10.1021/acs.analchem.0c03456.
[22] R.A. Fisher, Statistical methods for research workers, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in Statistics: Methodology and Distribution, Springer, New York, NY, 1992: pp. 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6.
[23] K.M. Pierce, L.F. Wood, B.W. Wright, R.E. Synovec, A comprehensive two-dimensional retention time alignment algorithm to enhance chemometric analysis of comprehensive two-dimensional separation data, Anal. Chem. 77 (2005) 7735–7743. https://doi.org/10.1021/ac0511142.
[24] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.
[25] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-

flight mass spectrometry (GC×GC–TOFMS) data using a null distribution approach, Anal. Chem. 87 (2015) 3812–3819. https://doi.org/10.1021/ac504472s.
[26] B.P. Tu, A. Kudlicki, M. Rowicka, S.L. McKnight, Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes, Science 310 (2005) 1152–1158. https://doi.org/10.1126/science.1120499.
[27] R.R. Klevecz, J. Bolen, G. Forrest, D.B. Murray, A genomewide oscillation in transcription gates DNA replication and cell cycle, PNAS 101 (2004) 1200–1205. https://doi.org/10.1073/pnas.0306490101.
[28] A.J. Burnetti, M. Aydin, N.E. Buchler, Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates, Mol. Biol. Cell 27 (2016) 64–74. https://doi.org/10.1091/mbc.E15-07-0454.
[29] B.P. Tu, R.E. Mohler, J.C. Liu, K.M. Dombek, E.T. Young, R.E. Synovec, S.L. McKnight, Cyclic changes in metabolic state during the life of a yeast cell, PNAS 104 (2007) 16886–16891. https://doi.org/10.1073/pnas.0708365104.
[30] R.E. Mohler, B.P. Tu, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography–time-of-flight-mass spectrometry data, J. Chromatogr. A 1186 (2008) 401–411. https://doi.org/10.1016/j.chroma.2007.10.063.
[31] R. Bro, PARAFAC tutorial and applications, Chem. Intel. Lab. Sys. 38 (1997) 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4.
[32] G.F.L. Pereira, E.E. Fileti, L.J.A. Siqueira, Comparing graphite and graphene oxide supercapacitors with a constant potential model, J. Phys. Chem. C 125 (2021) 2318–2326. https://doi.org/10.1021/acs.jpcc.0c10347.
[33] G.C. Pratt, C.Y. Wu, D. Bock, J.L. Adgate, G. Ramachandran, T.H. Stock, M. Morandi, K. Sexton, Comparing air dispersion model predictions with measured concentrations of VOCs in urban communities, Environ. Sci. Technol. 38 (2004) 1949–1959. https://doi.org/10.1021/es030638l.
[34] F. Urbina, K.M. Zorn, D. Brunner, S. Ekins, Comparing the Pfizer central nervous system multiparameter optimization calculator and a BBB machine learning model, ACS Chem. Neurosci. 12 (2021) 2247–2253. https://doi.org/10.1021/acschemneuro.1c00265.
[35] H.T.M. Heinrich, P.J. Bremer, C.J. Daughney, A.J. McQuillan, Acid−base titrations of functional groups on the surface of the thermophilic bacterium Anoxybacillus flavithermus: comparing a chemical equilibrium model with ATR-IR spectroscopic data, Langmuir 23 (2007) 2731–2740. https://doi.org/10.1021/la062401j.
[36] T. Kasukawa, M. Sugimoto, A. Hida, Y. Minami, M. Mori, S. Honma, K. Honma, K. Mishima, T. Soga, H.R. Ueda, Human blood metabolite timetable indicates internal body time, PNAS 109 (2012) 15036–15041. https://doi.org/10.1073/pnas.1207768109.
[37] Y. Izadmanesh, E. Garreta-Lara, J.B. Ghasemi, S. Lacorte, V. Matamoros, R. Tauler, Chemometric analysis of comprehensive two-dimensional gas chromatography–mass spectrometry metabolomics data, J. Chromatogr. A 1488 (2017) 113–125. https://doi.org/10.1016/j.chroma.2017.01.052.
[38] M. Brokl, L. Bishop, C.G. Wright, C. Liu, K. McAdam, J.-F. Focant, Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, J. Chromatogr. A 1370 (2014) 216–229. https://doi.org/10.1016/j.chroma.2014.10.057.

[39] R. Watanabe, A. Sugahara, H. Hagihara, J. Mizukado, H. Shinzawa, Three-way evolved gas analysis-mass spectrometry combined with principal component analysis (EGA-MS-PCA) to probe interfacial states between matrix and filler in poly(styrene-b-butadiene-b-styrene) (SBS) nanocomposites, Polym. Test. 101 (2021) 107300. https://doi.org/10.1016/j.polymertesting.2021.107300.
[40] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1644 (2021) 462092. https://doi.org/10.1016/j.chroma.2021.462092.
[41] S. Wang, T. Kind, P.L. Bremer, D.J. Tantillo, O. Fiehn, Quantum chemical prediction of electron ionization mass spectra of trimethylsilylated metabolites, Anal. Chem. 94 (2022) 1559–1566. https://doi.org/10.1021/acs.analchem.1c02838.

Chapter 6. Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data

This chapter is reproduced from Lina Mikaliunaite, Robert E. Synovec, "Simultaneous Discovery of Compounds Dominated by either Molding Kinetics or Geographical Region of Origin for Moisture Damaged Cacao Beans using Orthogonally Applied Tile-Based Fisher Ratio Analysis of GC×GC-TOFMS Data" *J. Chromatogr. A* 1730 (2024) 465093.

6.1 Introduction

Comprehensive two-dimensional (2D) gas chromatography (GC×GC) paired with time-of-flight mass spectrometry (TOFMS) is becoming widely used in various application areas such as food [1–3], fuels [4,5], health monitoring [6–8], among many others [9,10]. With the growing use of GC×GC-TOFMS, there is also a growing need for developing chemometric techniques, both targeted and nontargeted, to analyze these large and complex datasets [11,12]. Nontargeted chemometric techniques are extremely important for tracking biological processes and differences between samples, as usually these samples have hundreds and possibly thousands of compounds present in them [12]. The added difficulty with biological samples is that they inherently will have more variation (often 30% - 100% *RSD*). For this purpose, Fisher ratio (F-ratio) analysis [13,14] is a popular and effective approach, where the "standard" F-ratio [15] is statistically defined as the ratio of the between-class variance of analyte signal relative to the within-class variance [16], based upon two or more sample classes as defined by the experimental design. With the development of tile-based F-ratio analysis, which has the benefit of accounting for peak shifting between sample runs, and its commercialization as ChromaTOF Tile [17], this data analysis tool readily facilitates a more thorough examination of interesting

datasets both newly collected [2,18,19] as well as older datasets collected before development of tile-based F-ratio analysis [10].

While the initial research-level and commercial introduction of tile-based F-ratio analysis for GC×GC-TOFMS datasets implemented the "standard" F-ratio definition, recent reports have demonstrated the promise of using modified versions of the F-ratio ranking metric, such as control-normalized tile-based Fisher ratio analysis [20] and minimum variance optimized tile-based Fisher ratio analysis [21]. However, with the implementation of these variations of tile-based F-ratio analysis, sample classes can only be assigned one time, thus are limited to investigate only one chemically related process and/or characteristic that may be changing in the dataset. In contrast, if the chemical system under investigation inherently has two processes and/or characteristics potentially changing at once, usually that would require a generation of two hit lists, and would require mining of both, matching analytes in one hit list to analytes in the other hit list, which would be time-consuming fraught with shortcomings. Thus, these issues generally necessitate designing experiments to only have one process/characteristic that needs to be examined, eg., comparing a patient class to a control class in a medical study [22]. However, in some cases analysts might have experimental designs that include two (or more) chemically related processes and/or characteristics. An in-depth and comprehensive analysis of such a dataset would be well served by the development of a version of the tile-based F-ratio software that can simultaneously mine more than one process/characteristic.

Thus, the goal of this report is to address this challenge for tile-based F-ratio analysis using a GC×GC-TOFMS dataset collected for moisture damaged cacao beans [23]. Specifically, cacao beans from six geographical regions were analyzed once a day for six days following the initiation of moisture damage to trigger the molding process. There are two "extremes" to the

experimental sample class design: six time points for the molding kinetics versus the six geographical regions of origin from around the world, resulting in a 6 × 6 element signal array from the GC×GC-TOFMS data referred to as a *composite chemical fingerprint* (CCF) for each analyte. While from the literature we know that mold produces volatile compounds [23–25], there is also evidence that geographical origin of beans, be it cacao or coffee, has an impact on the volatile profile of the beans [26–28]. This scenario presents a unique opportunity to create a computational approach to simultaneously tease apart these two sample class structures using the tile-based F-ratio analysis software platform with suitable modifications.

To this end, the tile-based F-ratio analysis software was modified in-house to simultaneously calculate two F-ratios at the 2D location for each hit ($\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$) defined by the two sample class dimensions in the 6 × 6 CCF data structure on a *per-tile, per m/z basis*. The six kinetic time points were stored along the rows, while the six geographical regions were stored along the columns. Thus, the $\text{F-ratio}_{\text{kinetic}}$ was calculated treating the signals of the molding kinetic time points as classes and functionally treating the same signals at a given time point along the columns, i.e., the geographical regions, as "replicates" of each other. Similarly, the $\text{F-ratio}_{\text{region}}$ were calculated by treating the geographical regions as the classes with all time points along the rows as "replicates" of each other. Since these two F-ratios are calculated for every tile, the final hit list is generated using the higher of the two F-ratios, thus ranking each hit in terms of which of the two sample characteristics dominates, either the molding kinetics or geographical region of origin. Note, this approach provides a hit list in which the top of the hit list is comprised of analytes that are most strongly dominated by either molding kinetics or by geographical region of origin, however the final hit list contains all analytes discovered, even those that contain some combination of both sample characteristics. One would imagine that

analytes with a high F-ratio$_{kinetic}$ are likely to be species not initially in any of the bean samples, but then are generated by the molding process. Likewise, analytes with a high F-ratio$_{region}$ are effectively not impacted by the molding but they do inherently differ from one geographical region of origin to the next. The software approach introduced herein will provide an effective means to discover these two "orthogonal" sample characteristics and arrive at two sets of the most robust (statistically significant) biomarkers to distinguish these characteristics.

6.2 Experimental

6.2.1 Cacao bean samples and sample preparation

GC×GC-TOFMS chromatograms (Fig. 6.1.) of cacao beans from six geographical origins (Costa Rica, Ghana, Ivory Coast, Venezuela, Ecuador, and Panama) that were progressively damaged by moisture were analyzed [23]. A stock sample of raw beans from each origin was stored under cool and dry conditions to preserve the bean quality prior to assessing the impact of moisture damage. For each origin, 18 representative beans were taken and placed in a sealable bag to which 10 mL of filtered water was added. To ensure the total time with added moisture was consistent for each sample at the time of analysis, the addition of water was staggered by 1 hr for each origin to compensate for GC×GC-TOFMS analysis time. The beans were stored at room temperature and samples, i.e., three representative beans per sample, were analyzed for a total of six time points (1, 2, 3, 4, 5, 6 days) to assess the moisture damage process.

Head-space solid phase microextraction (HS-SPME) was applied for sampling. A 65 μm PDMS/DVB SPME fiber (Sigma-Aldrich, St. Louis, MO, USA) preconcentrated the headspace analytes above each cacao bean sample. The fiber was conditioned at 250 °C for 30 min prior to each extraction. At the given time (1, 2, 3, 4, 5, 6 days), three cacao beans were removed from the sealable bag of a given origin and sealed within a new 15 mL SPME vial. For

consistency, the six geographical region samples were analyzed in the same order that water was added. Each bean sample was heated in a water bath to 60 °C for 15 min, after which the SPME fiber was exposed to the headspace for 10 min. After being extracted, beans were not returned to the bag as the extraction alters the beans.

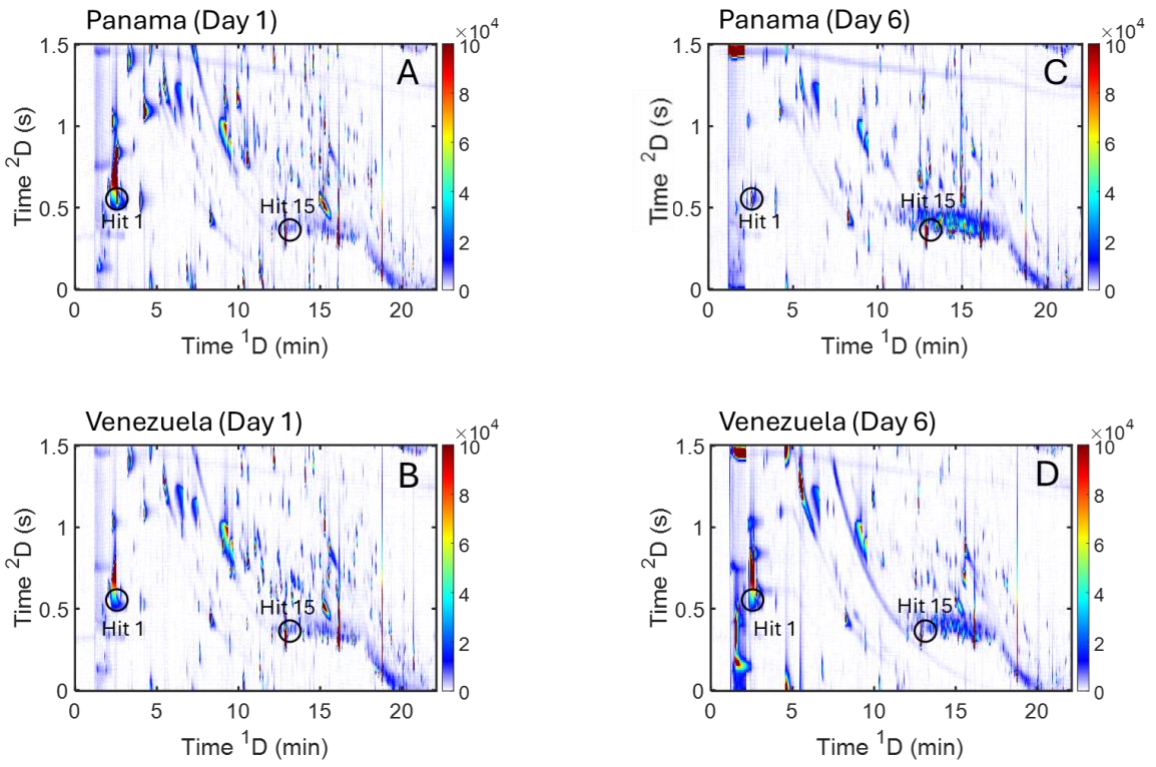


Figure 6.1. GC×GC total ion current (TIC) chromatograms of beans from two of the six geographical regions and from different molding days. Beans from (A) Panama and (B) Venezuela on Day 1 of the molding process. Then, beans from (C) Panama and (D) Venezuela on the Day 6 of molding. Chromatographic 2D locations of Hit 1 (3-methyl-2-nonene) and Hit 15 (2-methyl-3-buten-2-ol) are labeled on each chromatogram.

6.2.2 Chromatographic conditions

Following extraction, the SPME fiber was introduced to the GC×GC-TOFMS instrument comprised of an Agilent 6890N GC (Agilent Technologies, CA, USA) with a thermal modulator (4D upgrade, LECO, St. Joseph, MI, USA) and a Pegasus III TOFMS (LECO, St. Joseph, MI, USA). The inlet was maintained at 250 °C with a constant flow of helium at 1 mL/min, and the SPME fiber was kept in the inlet for 5 min for injection. The GC×GC column set consisted of a

non-polar ^{1}D column (20 m × 250 µm i.d. × 0.5 µm RTX-5MS film, Restek, PA, USA), and a polar ^{2}D column (2 m × 180 µm i.d. × 0.2 µm RTX-200MS film, Restek, PA, USA). For the first 5 min, during injection, the ^{1}D column was held at 40 °C and the ^{2}D column at 50 °C. Both columns followed a temperature program that ramped at a rate of 8 °C/min from 40 °C to 140 °C for the ^{1}D column with a +10 °C offset for the ^{2}D column. The rate was then increased to 30 °C/min to a final temperature of 250 °C for the ^{1}D column and 260 °C for the ^{2}D column, where the columns were held constant for an additional 1 min. The modulator temperature was maintained 20 °C higher than the ^{1}D column, with a modulation period of 1.5 s. The transfer line was held at 280 °C and the TOFMS ion source at 250 °C. Mass channels 40-250 *m/z* were collected and stored at a rate of 100 spectra/s.

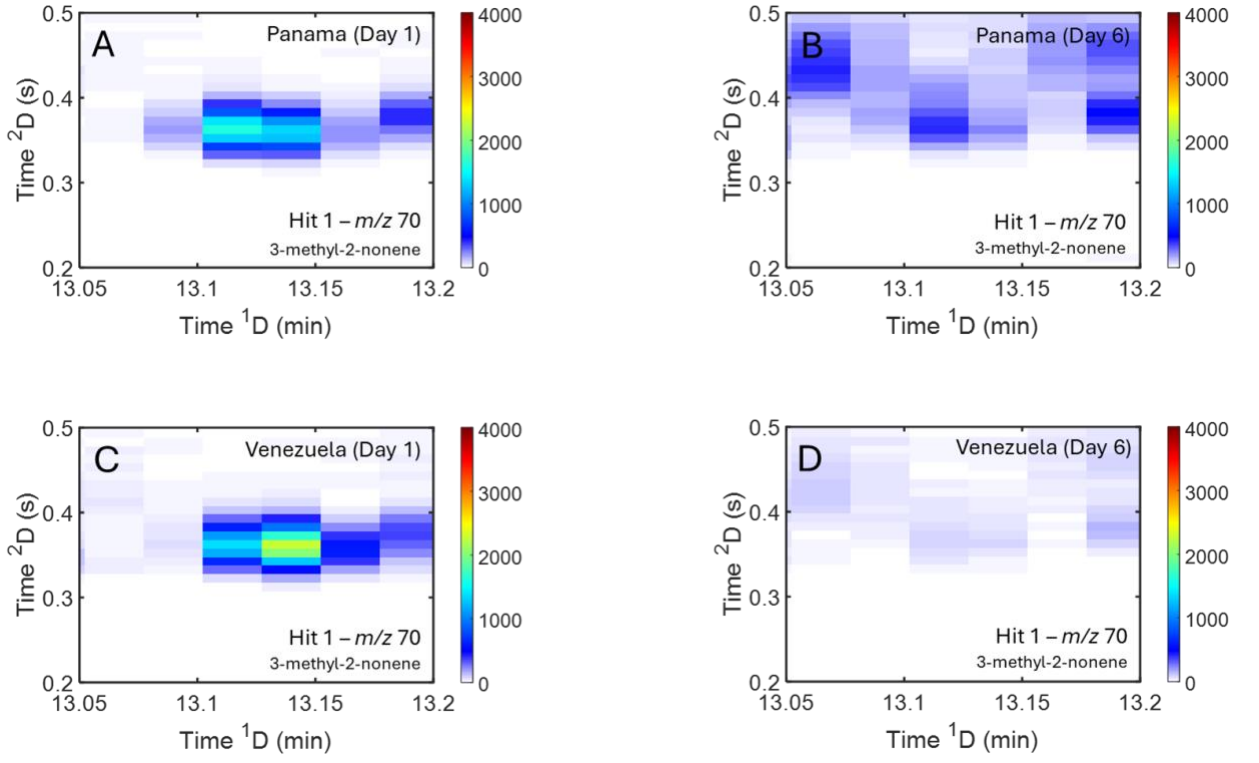


Figure 6.2. Selective ion chromatograms using the signal from m/z 70, which yielded the highest F-ratio for Hit 1 (3-methyl-2-nonene), a kinetic hit. For Day 1, 3-methyl-2-nonene has a high concentration both in (A) Panama and (B) Venezuela beans, however throughout the molding process, specifically by Day 6, 3-methyl-2-nonene almost completely disappears in both (C) Panama and (D) Venezuela beans.

6.2.3 Data Analysis

A total of 36 chromatograms tracking molding over the period of six days on beans of six different geographical regions (Panama, Costa Rica, Ivory Coast, Ghana, Ecuador, and Venezuela) were analyzed using our in-house tile-based F-ratio software [13] based upon Matlab 2023a (MathWorks, Inc., Natick, MA). Relative to previous reports, in this study the software code was modified to simultaneously discovery analytes impacted by molding kinetics and/or geographical region of origin, thus two separate F-ratios were calculated: $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$. Examples of discovered analytes are provided in Figs. 6.2.,6.3. For the $\text{F-ratio}_{\text{kinetic}}$ calculations, the six sequential days of molding kinetics were used as classes and bean

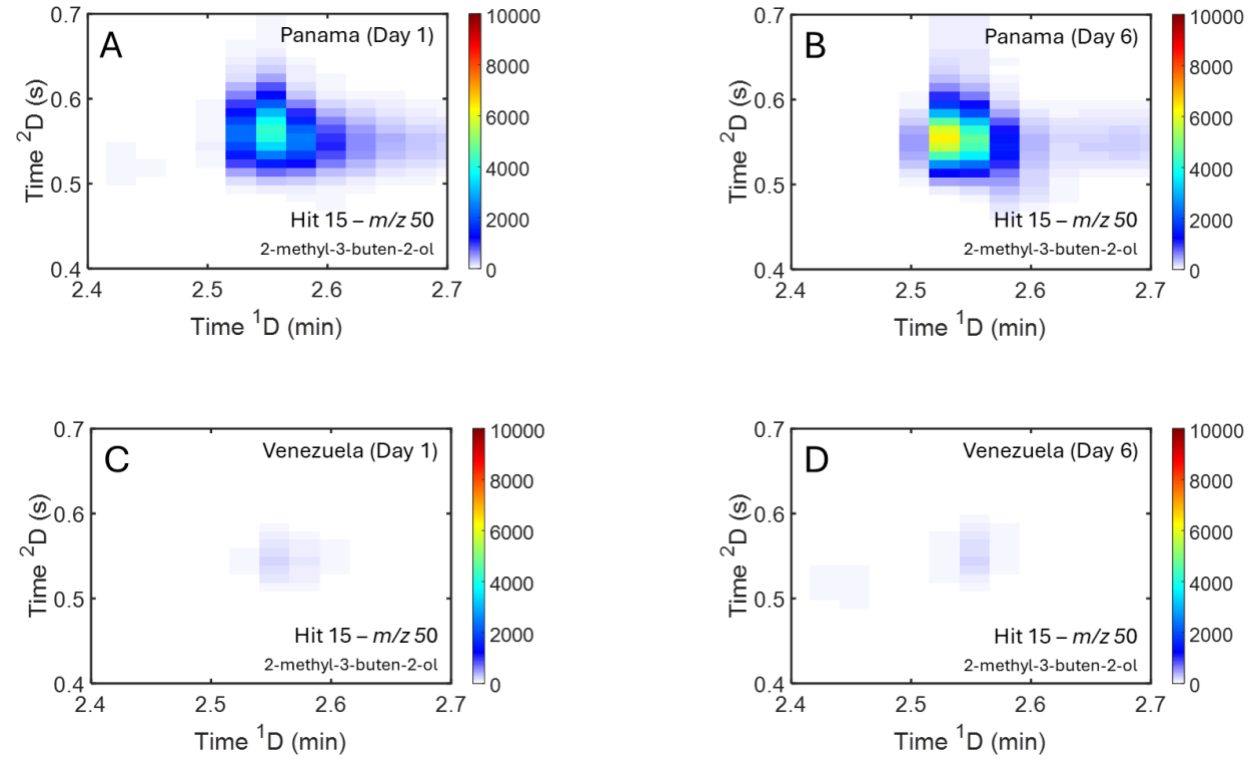


Figure 6.3. Selective ion chromatograms made using m/z 50, which yielded the highest F-ratio for Hit 15 (2-methyl-3-buten-2-ol), which is a geographical region hit. On Day 1, 2-methyl-3-buten-2-ol is seen to be present in Panama beans for (A) Day 1 and (B) Day 6, however, 2-methyl-3-buten-2-ol is not visually present in Venezuela beans for either (C) Day 1 (C) or (D) Day 6.

geographical regions of origin essentially served as "replicates" from the perspective of the F-ratio$_{kinetic}$ calculation. In contrast, for the F-ratio$_{region}$ calculations, the six geographical regions of origin defined the classes and the days of molding served as the "replicates." With tile-based F-ratio analysis, the pixel-level GC×GC data was baseline corrected and summed using four strategically overlapped 2D grids of 2D tiles, as a function of mass channel (*m/z*) which provides data reduction along both separation dimensions [13,14]. A tile size of 7.5 s (5 modulations) on ^{1}D × 150 ms on ^{2}D was implemented to encompass the peak widths in both dimensions and the modest run-to-tun retention shifting observed between sample runs. Since there were 100 spectra/s collected with a mass range of *m/z* 40-250, each tile contained 75 pixels from the initial dataset. A signal-to-noise (*S/N*) threshold of 10 of the tiled signal data was applied to exclude *m/z* per tile having low *S/N*. This *S/N* threshold was set to a signal equal to three times the standard deviation (3s) of a tiled noise region, which was taken as the first 1 min of detected signal, during which no peaks eluted, from a representative chromatographic run. The two separate F-ratios for each 2D tile (F-ratio$_{kinetic}$ and F-ratio$_{region}$) were then calculated for each *m/z* clearing the *S/N* threshold at least in one replicate of either class. This ensures the F-ratios are being calculated by summing the signal in the same tile per chromatogram for both F-ratio$_{kinetic}$ and F-ratio$_{region}$ guaranteeing an objective comparison. Since there are four overlapping tile grids, which are necessary to optimally discover all analytes while mitigating retention shifting, the redundant hits for both types of F-ratios initially obtained were removed using a "pin and cluster" algorithm [14], with the cluster tile size identical to the 2D grid tile dimensions of 7.5 s (5 modulations) on ^{1}D × 150 ms on ^{2}D. Finally, the highest of each of the two F-ratios (F-ratio$_{kinetic}$ and F-ratio$_{region}$), per hit tile at their respective *m/z* channel that produced highest F-ratio, were stored for every analyte hit. However, the *higher* of the two F-ratios was used to rank

the analytes in the final hit list, along with an indicator column specifying which of the two F-ratios dominated. Additionally, all siloxane related peaks were removed as artifacts of the initial SPME sampling leading to the GC×GC-TOFMS data collection. Once the final hit list was established, the next step was to examine more broadly the signal patterns that were observed for each analyte in the hit list. To this end, for all analytes in the hit list their signal "pattern" from the six geographical regions of origin collected over six days of molding kinetics was obtained using the tile centered on the analyte pin location using the *m/z* that produced the highest F-ratio to arrive at a 6 × 6 signal element CCF signal matrix array (Hits 1 and 15 in Fig. 6.4. A-B).

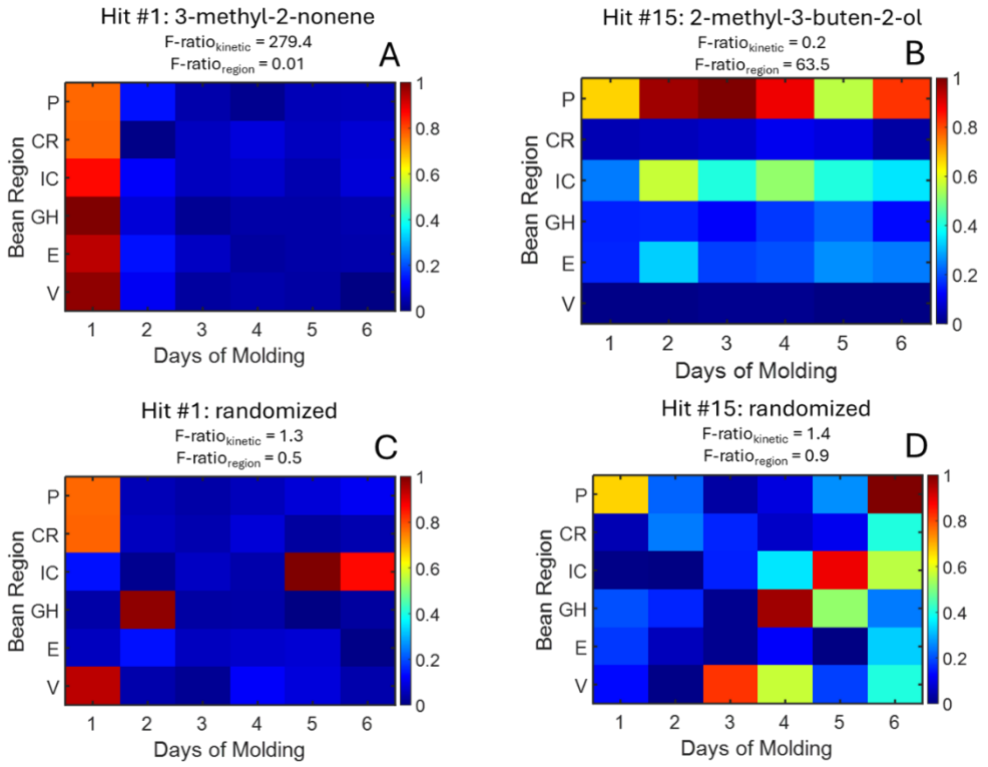


Figure 6.4. Composite chemical fingerprints (CCFs) for (A) 3-methyl-2-nonene (Hit 1) and (B) 2-methyl-3-buten-2-ol (Hit 15). The 36 signals in each CCF have been scaled to the highest signal, where dark red represents the highest relative signal of 1, and dark blue represents the lowest relative signal of 0. Bean type labels: Panama (P), Costa Rica (CR), Ivory Coast (IC), Ghana (G), Ecuador (E), and Venezuela (V). (A) 3-methyl-2-nonene has high signal only for Day 1, producing an F-ratio$_{kinetic}$ of 279.4 and F-ratio$_{region}$ of 0.01. (B) 2-methyl-3-buten-2-ol has relatively high signals for all of the molding days in the beans from the Panama region as well as Ivory coast, resulting in an F-ratio$_{region}$ of 63.5 and an F-ratio$_{kinetic}$ of 0.2. (C) An example of when the signals in these CCFs are randomized for 3-methyl-2-nonene, producing a null CCF pattern resulting in an F-ratio$_{kinetic}$ of 1.3 and F-ratio$_{region}$ of 0.5. (D) Similarly, an example of randomizing the signals for 2-methyl-3-buen-2-ol, producing a null CCF pattern resulting in an F-ratio$_{kinetic}$ of 1.4 and F-ratio$_{region}$ of 0.9.

For this study, we needed to determine which analyte hits possessed "statistically significant" information in the context of expressing primarily molding kinetics or geographical region of origin. To do so, we used pseudo-null distributions to define significance thresholds of the F-ratio hit list generated as presented herein. The pseudo-null distributions approach is based upon using a representative fraction of all possible signal combinations in the CCF of each discovered analyte and randomizing them to find an F-ratio threshold which excludes hits that do not exhibit sufficient statistical significance [14]. However, if all possible random combinations of each CCF were used, there would be 36! combinations (since there are 36 signals in each analyte CCF), equal to 3.71×10^{41} combinations for every hit. Since that number of calculations is computationally excessive, we elected to use 5000 of the unique combinations for every hit to serve as a reasonable sampling of the randomness. Thus, for each analyte hit with its CCF, there were two F-ratios, $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$ (ranked by the higher of the two), followed by 5000 unique randomized CCFs, which in turn produced a pair of "null" F-ratios: null $\text{F-ratio}_{\text{kinetic}}$ and null $\text{F-ratio}_{\text{region}}$. Examples of randomizing the signals in the CCF for two hits (Hit 1 and Hit 15) are provided in Fig. 6.4. C-D, with corresponding null F-ratio values. We shall see there are 591 analyte hits in the hit list, resulting in 5000 × 591, or 2.955×10^{6} null CCFs.

Next, the F-ratios for all of these 2.955×10^{6} null CCFs were calculated and used to create a pseudo-null distribution for both $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$, resulting in 2.955×10^{6} for each kinetic and region F-ratio calculations (for a total of 5.91×10^{6} F-ratios). A composite null distribution of all 5.91×10^{6} F-ratios was applied since the kinetic and region pseudo-null distributions were essentially identical. Statistical thresholds varying from 99.0% to 99.999% were explored with the composite pseudo-null distribution in the context of the experimentally obtained $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$ distributions. By this approach we sought to obtain a better

understanding of what null threshold corresponded to visually recognizable patterns in the experimentally obtained CCFs that are clearly dominated by either molding kinetics or geographical region of origin. For this purpose, classification of all the analyte CCFs passing the 99.9% null distribution threshold was performed using PCA with the PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA). Finally, the most *kinetic-versus-region* distinguishing hits using a 99.999% null distribution threshold were identified using PARAFAC via a three-way array (retention × spectra × samples). The purified spectra were then matched to the NIST library and identified if they had a match value (MV) of 800 or above.

6.3 Results and Discussion

Chromatograms tracking molding over 6 days of beans from 6 geographical regions (Panama, Costa Rica, Ivory Coast, Ghana, Ecuador and Venezuela) were analyzed using a new method for simultaneous discovery of analytes using the modified tile-based F-ratio tile code. While previously the data was collected to track analytes most obviously impacted only by molding [23], in the present study we focus also on the fact that concurrent with the molding effect there is also information regarding the geographical region that is inherently present in this dataset. In order to frame these two characteristics of the cacao beans, we selected chromatograms from two of the six regions (Panama versus Venezuela) from two days (Day 1 versus Day 6) in Fig. 6.1. While the chromatograms in Fig. 6.1. appear rich in chemical information, the complexity of the entire chromatograms does not readily afford visual inspection to discover all of the analytes changing due to regarding their molding kinetics and geographical region characteristics. At this stage, we provide examples of analytes subsequently discovered by tile-based F-ratio analysis labeled as Hit 1 (3-methyl-2-nonene), a kinetic dominated hit, and Hit 15 (2-methyl-3-buten-2-ol), a geographical region dominated hit.

Specifically for 3-methyl-2-nonene in Fig. 6.2., this analyte is present in Day 1 for both Panama and Venezuela (Figs. 2A,C), but then the signal is significantly lower by Day 6 (Fig. 6.2. B,D), thus is an excellent example of a kinetic hit. In contrast, analytes that are observed to be robust to the molding effect can maintain and thus express their inherent differences in geographical origin. This is the case for 2-methyl-3-buten-2-ol in Fig. 3, which is observed at a much higher concentration in the Panama region sample (Fig. 6.3. A-B), relative to the Venezuela region sample (Fig. 6.3. C-D), while the analyte concentration is essentially unchanged from Day 1 to Day 6. We want to reinforce the reality that neither of these two analytes in Figs. 6.2.,6.3. that express these two sample class dependencies (molding kinetics and the geographical region of origin) can be readily found by visual inspection. Indeed, these two example analytes (which were found by the subsequent F-ratio analysis) are instructive to highlight both the benefit and challenge of finding all analytes that are changing due to molding kinetics and/or geographical region. To comprehensively address this challenge, we have implemented tile-based F-ratio analysis using the algorithm modified to simultaneously discover both sample class dependencies.

With the previously reported tile-based F-ratio software (either our in-house code or the commercial options), only one of these two sample class dependencies could be analyzed at a given time, thus generating two separate hit lists. Then, the analyst would need to match analyte peaks across both hit lists. This would also be time consuming and prone to many mistakes, and so to address these issues, the F-ratio tile code was modified for this current study to simultaneously calculate two separate F-ratios for each tile: $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$. Thus, a single hit list is generated with the analyte hits ranked using higher F-ratio of the two ($\text{F-ratio}_{\text{kinetic}}$ or $\text{F-ratio}_{\text{region}}$) whereby the magnitude of the F-ratio is used to rank the hit list. The hit

list, after the removal of redundant hits and artifact hits such as siloxanes generated by SPME, contained 591 hits. The composite chemical fingerprint (CCF) for each of the 591 hits was also obtained. For example, the signals for Hit 1 (3-methyl-2-nonene) in Fig. 2 are merely the summed signal of all pixels at *m/z* 70, for the Days 1 and 6 for both Panama and Venezuela. The CCF for the highest ranked molding kinetic hit, 3-methyl-2-nonene (Hit 1), produced an F-ratio$_{kinetic}$ of 279.4 (Fig. 6.4. A), with Days of Molding expressing clearly observed sample classes. This analyte is present at a higher concentration on Day 1 but drops very quickly to a much lower concentration by Day 2 mark and stays at a low concentration through Day 6. Concurrently, the molding kinetic pattern for 3-methyl-2-nonene is very consistent from one geographical region to the next, thus producing a very low F-ratio$_{region}$ of 0.01. In contrast, the CCF for a highly ranked analyte that was dominated by possessing differences due to geographical region of origin was 2-methyl-3-buten-2-ol (Hit 15) with an F-ratio$_{region}$ of 63.5 using *m/z* 50 (Fig. 4B), with Bean Region expressing clearly observed sample classes, where 2-methyl-3-buten-2-ol can be seen present for Panama and Ivory Coast beans, and to the smaller extent for Ecuador beans, but not the other three regions. While the CCF pattern for 2-methyl-3-buten-2-ol was dominated by geographical region of origin, concurrently this analyte expressed little to no change due to molding kinetics and thus produced an F-ratio$_{kinetic}$ of 0.2. In conjunction with the hit discovery by the modified F-ratio software, the signal data for all 591 analyte hits were processed to produce their individual CCF as in Fig. 6.4. A-B, and subsequently related to each other via PCA, *vide infra*.

While the previous discussion of Hits 1 and 15 provided specific examples, next we summarize the F-ratio distributions for all 591 hits (kinetic versus region) and relate them to the composite pseudo-null F-ratio distribution. With 591 hits in the F-ratio ranked hit list, one could

imagine that the hits near the top of the hit list would likely be dominated by either molding kinetics (Fig. 6.5. A) or geographical region of origin (Fig. 6.5. B), and that further down the hit list that many of the analytes would be more inclined to possess relatively even and/or seemingly random contributions of these two sample class dependencies. We know that the F-ratio software is designed to find all of the analyte peaks that possess sufficient signal to surpass the *S/N* threshold and then to simply calculate the F-ratio of each hit. Thus, we expect that at the bottom of the hit list that the CCFs might look more like a random pattern of signals. The notion being that any CCF that possessed a random pattern could not be classified as being sufficiently dominated by either molding kinetics or geographical region of origin. To determine a "statistically significant" threshold for this classification exercise, we used pseudo-null distributions to define significance thresholds of the F-ratio hit list generated. As explained in the Experimental section, 5000 randomly generated null CCFs were prepared from each of the 591 experimentally obtained CCF on a per analyte hit basis. While the experimentally obtained CCFs for Hits 1 and 15 (3-methyl-2-nonene and 2-methyl-3-buten-2-ol, respectively) are shown in Fig. 6.4. A-B, examples of randomizing their CCF patterns are provided for in Fig. 6.4. C-D, respectively. As anticipated for the purpose of producing a null CCF, the randomization of the signals from the experimental CCFs leads to very low F-ratios. For 3-methyl-2-nonene (Fig. 4C), the null F-ratio$_{kinetic}$ is only 1.3 while the null F-ratio$_{region}$ is only 0.5. Similarly, for 2-methyl-3-buten2-ol (Fig. 6.4. D), the null F-ratio$_{kinetic}$ is 1.4 and the null F-ratio$_{region}$ is 0.9. Overall, for the 591 hits there were 2.955×10^6 randomized CCFs generated, which in turn were used to calculate their null F-ratios for molding kinetics and region of origin. Using the null F-ratio$_{kinetic}$ and F-ratio$_{region}$ (2.955×10^6 of each type of F-ratio) of all of the randomized CCFs, pseudo-null F-ratio distributions for kinetics and region were generated. Since both of these null F-ratios

distributions were functionally equivalent, a single composite null F-ratio distribution was prepared (Fig. 6.5. C) to relate to the experimentally obtained F-ratio$_{kinetic}$ distribution (Fig. 6.5. A) and the F-ratio$_{region}$ distribution (Fig. 6.5. B). It is evident that the null F-ratio distribution lies to the left of the F-ratio$_{kinetic}$ and F-ratio$_{region}$ distributions. This suggests that most of the analyte hits do contain some level of non-random chemical and/or biological information present in their CCF signal patterns. For a more rigorous inspection of the various analytes in the hit list, we can calculate different thresholds using the null F-ratio distribution in Fig. 6.5. C. We chose to

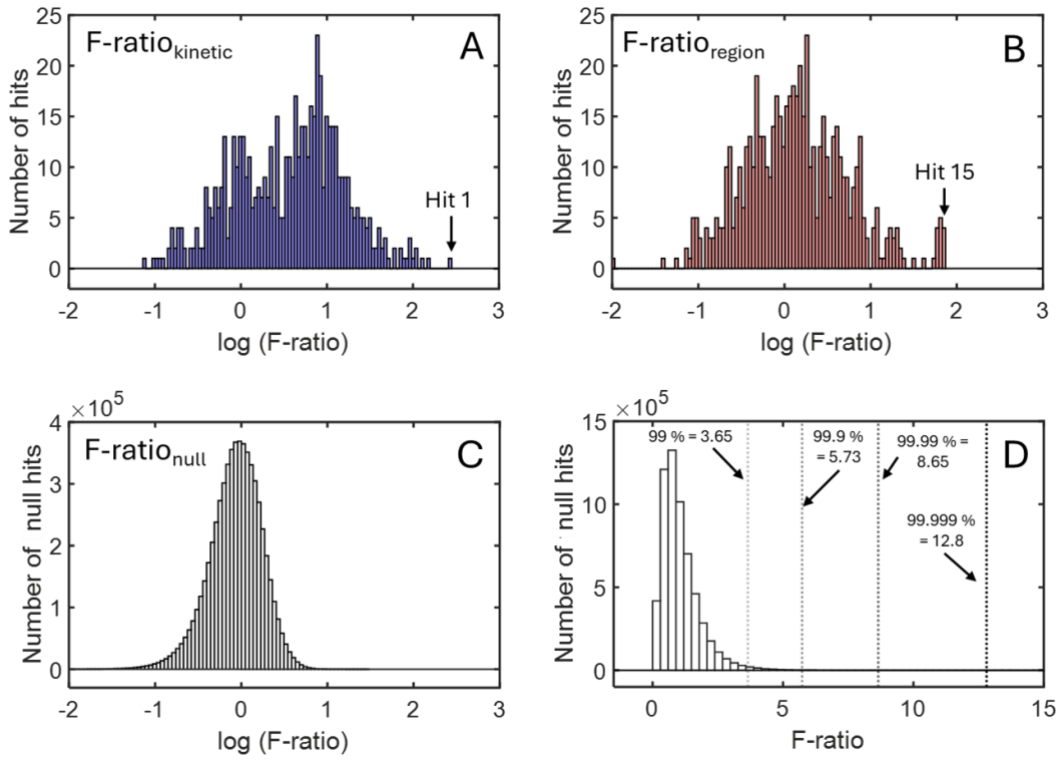


Figure 6.5. F-ratio distributions for (A) F-ratio$_{kinetic}$ and (B) F-ratio$_{region}$ values using log scale. (C) Distribution for F-ratio values in log scale for the 5.9×10^6 randomized null CCFs. (D) Various percentile thresholds for the pseudo-null distribution at 99%, 99.9%, 99.99% and 99.999% using a zoomed-in distribution for F-ratio values in log scale for the randomized null CCFs.

investigate thresholds of 99 %, 99.9 %, 99.99 % and 99.999 %, that resulted in the null F-ratio of these thresholds being 3.65, 5.73, 8.65, and 12.8, respectively. The number of hits that passed these F-ratio thresholds were 458, 339, 218 and 127, respectively. The 99% threshold still allows

for the majority of the hits (458 out of the original 591) to be deemed significant, in that all hits meeting or exceeding this threshold expressed at least a modest level of either kinetic or geographical region signal pattern in their CCF. While at the other extreme, 127 hits passing the 99.999% null threshold clearly expressed either a dominating kinetic or geographical region CCF pattern; examples of such patterns can be seen in Fig. 6.4. A-B for Hit 1 (dominating kinetic pattern) and Hit 15 (dominating region pattern).

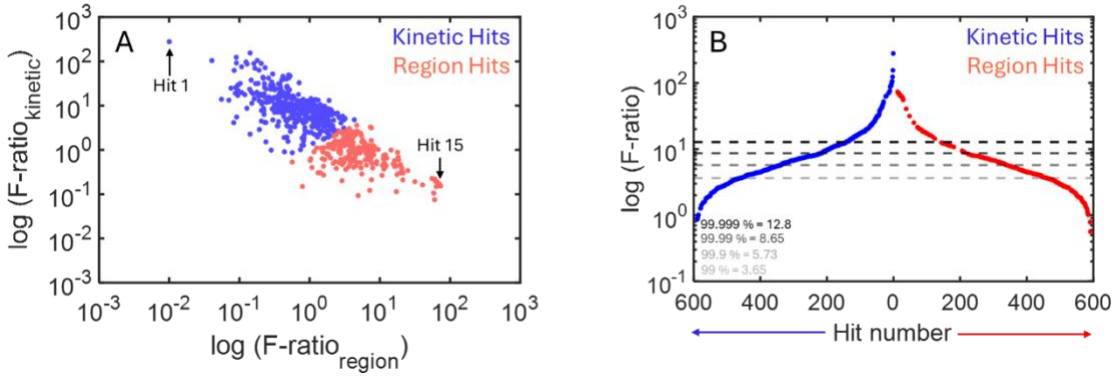


Figure 6.6. (A) Relationship between F-ratio$_{kinetic}$ and F-ratio$_{region}$ for every hit presented in log scale. Hits with a higher F-ratio$_{kinetic}$ are in blue and the hits with a higher F-ratio$_{region}$ are in light red. (B) Relationship between the highest F-ratio values from (A) in log scale and hit number. Various thresholds are marked with a dashed line, with the darkest color denoting the strictest threshold of 99.999%, and the lightest color the threshold of 99%.

To gain a better understanding of these thresholds we must look at the relationship of F-ratio$_{kinetic}$ and F-ratio$_{region}$ values as well as hit number (Fig. 6.6.). In Fig. 6.6. A, the F-ratio$_{kinetic}$ and F-ratio$_{region}$ coordinates for every hit are plotted as a single point using in a log scale. The general trend is that when the F-ratio$_{kinetic}$ is relatively high, the F-ratio$_{region}$ is relatively low, and vis versa, when the F-ratio$_{region}$ is high then the F-ratio$_{kinetic}$ is low (essentially an inverse relationship). Good examples of this are Hits 1 and 15 (highlighted in Figs. 6.2.-6.4.), that are indicated in Fig. 6.6. A. Another related relationship that needs to be explored is between F-ratio values (in log scale) to the hit number in the hit list (Fig. 6.6 B). In this plot only the maximum F-ratio is plotted per hit, whether it is the F-ratio$_{kinetic}$ or F-ratio$_{region}$. This further shows where

these different thresholds (obtained from Fig. 6.5 D) are compared to F-ratio values as well as hit number, indicated by dashed lines in Fig. 6.6. B. The CCFs of hits that are below 99% threshold appear "random" to visual inspection. Again, here we can see that the 99% threshold allows for the majority of the hits (458 out of the original 591) to be deemed significant. While the 99.999% null threshold allows a sizable portion of hits (127 out of the original 591) to pass, with a concomitant clear expression of either a kinetic or geographical region CCF pattern.

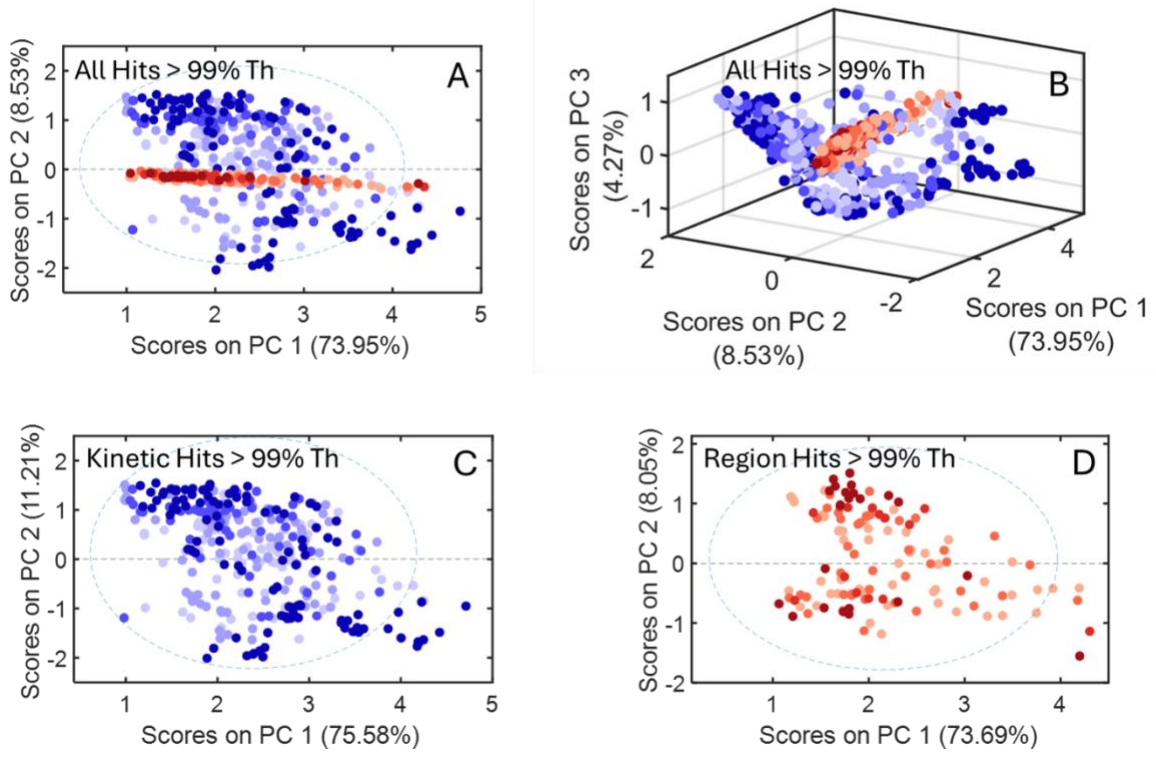


Figure 6.7. PCA scores plots for (A) all the hits exceeding the 99% null threshold (Th) using 2 principal components (PCs), (B) when using 3 PCs. PCA scores plots for (C) only the kinetic hits exceeding the 99% null threshold and (D) only the region hits exceeding the 99% null threshold. Kinetic hits are in blue shaded circles and region hits are in red shaded circles. The darkness of color shade represents what null distribution threshold the hit was found 99.999% being represented by the darkest color shade and 99% by the lightest color shade, with 99.99% and 99.9% being one hue lighter and two hues lighter, respectively.

Next, the CCF signal patterns for all hits at or above the 99% threshold were subjected to PCA, with various plots provided in Fig. 6.7. The different shades of color blue (kinetic hits) and red (region hits) indicate which F-ratio threshold was used to deem the hit significant. The

darkest color indicates the strictest threshold of 99.999%, and then the next shade lighter is 99.99%, with the next lighter at 99.9%, and finally the lightest shade exceeds the 99% threshold. In Figs. 6.7. A-B, we see that the kinetic and region hits separate from one another into two "classes" in PC scores space. Further, we also applied PCA separately to the kinetic and region hits, which are shown in Fig. 6.7. C and Fig. 6.7. D, respectively. While the hits are clearly separated in the PCA scores space, there are not very clear "sub-classes", so to better visualized what is going on in the PCA scores space, we must investigate how the CCF signal patterns relate to the position of each hit in the scores space.

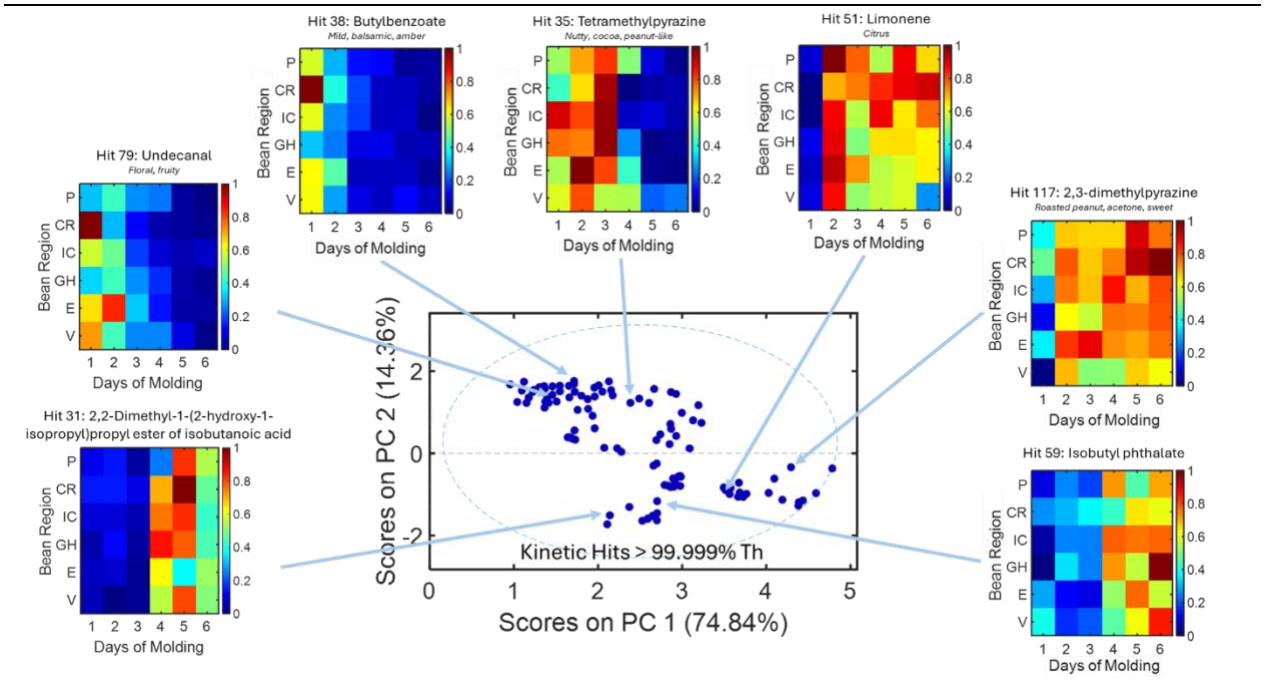


Figure 6.8. PCA scores plot of the kinetic hits that passed the strictest 99.999% null threshold. Also indicated are the color-coded composite chemical fingerprints (CCFs) for seven example analyte hits (as in Fig. 6.4).

To this end, we implemented PCA by focusing only on the 127 hits that cleared the 99.999% null F-ratio threshold. The scores plot of the kinetic hits passing this threshold is in Fig. 8, while the scores plot for the region hits is provided in Fig. 6.9. These are the same hits represented by the most intense dots in Fig. 6.7. Examples of the CCF signal patterns are also

provided in Figs. 6.8 and 6.9. For the kinetic hits in Fig. 6.8, it appears that PC1 separates hits roughly based upon the kinetic rate of concentration change, with the more rapidly changing analytes on the left side of PC1, and the more slowly changing analytes on the right side of PC1. Additionally, we note that the CCF patterns exhibit either a concentration decrease with time on the top side of PC2, or an increase with time on the bottom side of PC2. Likewise, some examples of the geographical region hits that passed a 99.999% threshold for null distribution are shown in Fig. 6.9. Here we can see clearer grouping of the hits into clusters, where the analytes that have positive PC2 are all in the cluster and seem to have high signal in the Costa Rica beans. The analytes with negative scores have low signal for Costa Rica beans, but high signals for Panama and/or Ecuador beans.

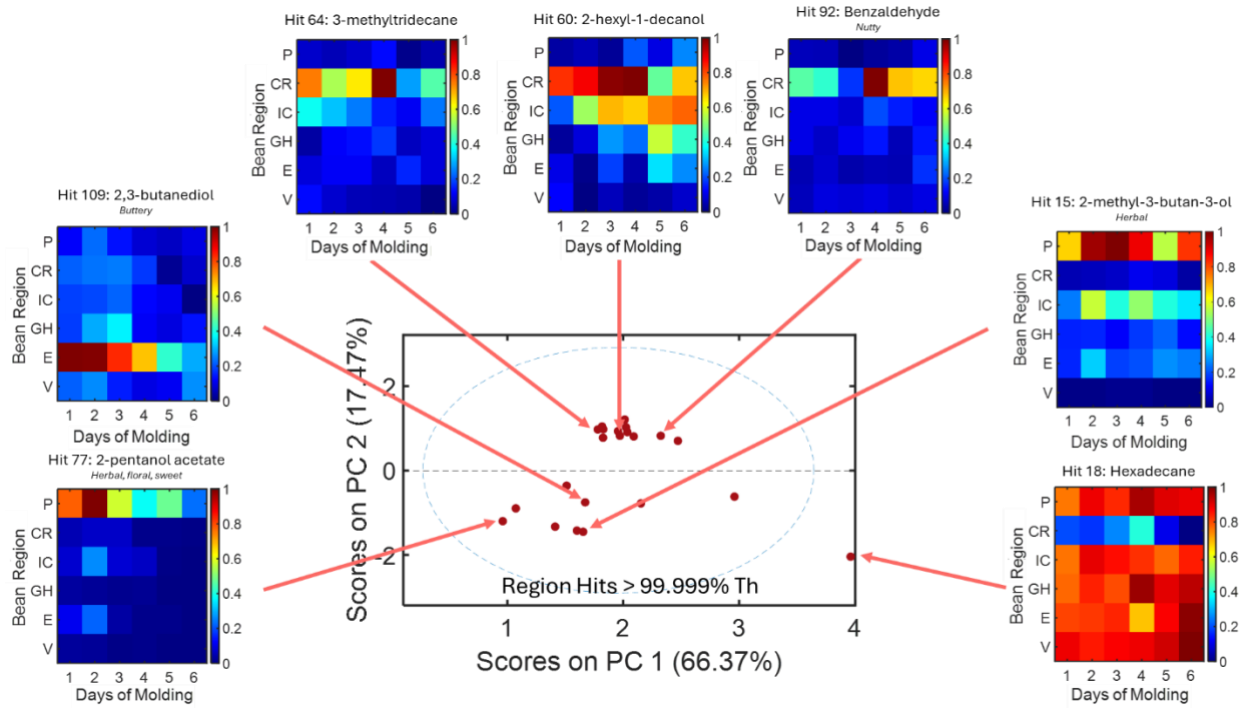


Figure 6.9. PCA scores plot of the region hits that passed the strictest 99.999% null threshold. Also indicated are the CCFs for seven example analytes.

Further discussion of the top 127 analytes passing the 99.999% null threshold is warranted, as many of them possess important properties related to the quality of the cacao

beans. These analytes were deconvoluted using PARAFAC, with the PARAFAC output providing a purified mass spectrum to optimize matching to the NIST library. A match value (MV) of 800 or above was deemed an adequate match for identification for the purposes of this study. For example, undecanal (MV: 924) exhibits a kinetic decrease (Hit 79 in Fig. 6.8.) and has been described to have floral and fruity flavors [29]. Hit 38 of butylbenzoate (MV: 822) has been described to be mild, balsamic, and amber flavor [30]. Another example is of an analyte that decreases kinetically is tetramethylpyrazine (MV: 868) (Hit 35 in Fig. 6.8.), which presents nutty, cocoa and peanut-like flavors [31] and has been found in cocoa powder [32]. There are also some good examples of analytes that increase in concentration with days of molding. One such example is diisobutyl phthalate (MV: 859) (Hit 59 in Fig. 6.8.), which is a known toxin, and has been shown to have adverse health effects in rats [33]. However, the increase of diisobutyl phthalate could also be coming not from the mold but rather from the plastic bag that the beans were stored in during molding [34]. Further, 2,3-dimethylpyrazine (MV: 924) increases with concentration (Hit 117 in Fig. 6.8.) and expresses a roasted peanut, acetone, and sweet odor [35]. Finally, another example of a kinetic hit related to its properties is limonene (MV: 914) (Hit 51 in Fig. 6.8.), which is known to have citrus flavor, but also was shown to be related to molding [36]. Now, we briefly discuss a few examples of analytes that hold steadfast in their concentration, impervious to the molding effect, and thus are robust to their geographical region. Panama beans have a high concentration of 2-pentanol acetate (MV: 939) (Hit 77 in Fig. 6.9.), which is known to have herbal, floral and sweet flavor and has been found in bananas [37]. Another analyte that had a high concentration in Panama beans is 2-methyl-3-butan-3-ol (MV: 886) (Hit 15 in Fig. 6.9.), which has green herbal aroma and has been found in panned oolong tea [38]. Similarly, but in Ecuador beans, 2,3-butanediol (MV: 865) (Hit 109 in Fig. 6.9.) has a

buttery aroma [39] and has been found in the processing of cacao beans [40]. In Costa Rica beans, 3-methyltridecane (MV: 885) (Hit 64 in Fig. 6.9.), and benzaldehyde (MV: 900) (Hit 92 in Fig. 6.9.) have been found to have high concentrations. While a flavor profile for 3-methyltridecane could not be tracked down, it has been shown to exist in various foods [41,42]. Benzaldehyde has a nutty flavor [43] and has been found in cacao in previous studies [44]. It also has been noted to have anti-molding effects, which might provide insight into why it was not affected by molding of cacao beans [45].

6.4 Conclusion

Herein, we present a new computational method for simultaneous discovery of compounds indicating either molding kinetics or bean geographical region in moisture damaged cacao beans. The dataset of moisture damaged cacao beans where molding of six different bean types was monitored over six days was chosen for the method development, as this presented an opportunity to investigate two processes that were present inherently in the dataset at once: molding kinetics and bean type based on geographical region. The new method simultaneously calculates two F-ratios per analyte hit: one for $\text{F-ratio}_{\text{kinetic}}$ and $\text{F-ratio}_{\text{region}}$ and ranks the hit list using the higher of the two F-ratios. This method could be applied in principle to various studies in which class membership can be assigned in more than one dimension based upon the experimental design. While in this study equal numbers of classes were examined, six kinetic time points versus six geographical regions, implementation of the methodology does not require equal membership to both classes. For this study, null distributions were used to find different statistical thresholds to further categorize the discovered analytes. Using this new computational method presented here, we evaluated a moisture damaged cacao beans dataset and found 591 compounds that contribute to either molding kinetics or geographical region of origin. While 458

of the initial 591 compounds on the hit list passed the 99% null threshold indicating they possessed a “reasonable” amount of statistically significant chemical information to belong to one of these two classes, the CCF signal pattern of most of these analytes could not be readily interpreted by the human eye. However, at the 99.999% null threshold, all 127 analytes expressed a CCF signal pattern that was readily interpreted by the human eye as being either kinetic dominated or region dominated.

6.5 References

[1] C.N. Cain, M. Gaida, P.-H. Stefanuto, J.-F. Focant, R.E. Synovec, S.C. Jackels, K.J. Skogerboe, Investigating sensory-classified roasted arabica coffee with GC × GC-TOFMS and chemometrics to understand potato taste defect, Microchem. J. 196 (2024) 109578. https://doi.org/10.1016/j.microc.2023.109578.

[2] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A 1662 (2022) 462735. https://doi.org/10.1016/j.chroma.2021.462735.

[3] M.C. Rosso, M. Mazzucotelli, C. Bicchi, M. Charron, F. Manini, R. Menta, M. Fontana, S.E. Reichenbach, C. Cordero, Adding extra-dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential, J. Chromatogr. A 1614 (2020) 460739. https://doi.org/10.1016/j.chroma.2019.460739.

[4] G.S. Ochoa, R.E. Synovec, Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, Talanta 259 (2023) 124525. https://doi.org/10.1016/j.talanta.2023.124525.

[5] T.J. Trinklein, J. Jiang, R.E. Synovec, Profiling olefins in gasoline by bromination using GC×GC-TOFMS followed by discovery-based comparative analysis, Anal. Chem. 94 (2022) 9407–9414. https://doi.org/10.1021/acs.analchem.2c01549.

[6] P.C.K. Makhubela, E.R. Rohwer, Y. Naudé, Detection of tuberculosis-associated compounds from human skin by GC×GC-TOFMS, J. Chromatogr. B 1231 (2023) 123937. https://doi.org/10.1016/j.jchromb.2023.123937.

[7] R. Pesesse, P.-H. Stefanuto, F. Schleich, R. Louis, J.-F. Focant, Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC×GC-TOFMS, J. Chromatogr. B 1114–1115 (2019) 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029.

[8] A.Z. Berna, E.H. Akaho, R.M. Harris, M. Congdon, E. Korn, S. Neher, M. M’Farrej, J. Burns, A.R. Odom John, Reproducible breath metabolite changes in children with SARS-

CoV-2 infection, ACS Infect. Dis. 7 (2021) 2596–2603. https://doi.org/10.1021/acsinfecdis.1c00248.
[9] J. Kaur, N. Sun, J.E. Hill, Comprehensive profiling of terpenes and terpenoids in different cannabis strains using GC×GC-TOFMS, Separations 10 (2023) 500. https://doi.org/10.3390/separations10090500.
[10] L. Mikaliunaite, R.E. Synovec, Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, Talanta 244 (2022) 123396. https://doi.org/10.1016/j.talanta.2022.123396.
[11] P.-H. Stefanuto, A. Smolinska, J.-F. Focant, Advanced chemometric and data handling tools for GC×GC-TOF-MS: Application of chemometrics and related advanced data handling in chemical separations, TrAC, Trends Anal. Chem. 139 (2021) 116251. https://doi.org/10.1016/j.trac.2021.116251.
[12] T.J. Trinklein, C.N. Cain, G.S. Ochoa, S. Schöneich, L. Mikaliunaite, R.E. Synovec, Recent advances in GC×GC and chemometrics to address emerging challenges in nontargeted analysis, Anal. Chem. 95 (2023) 264–286. https://doi.org/10.1021/acs.analchem.2c04235.
[13] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.
[14] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC–TOFMS) data using a null distribution approach, Anal. Chem. 87 (2015) 3812–3819. https://doi.org/10.1021/ac504472s.
[15] R.A. Fisher, Statistical methods for research workers, in: S. Kotz, N.L. Johnson (Eds.), Breakthroughs in statistics: methodology and distribution, Springer, New York, NY, 1992: pp. 66–70. https://doi.org/10.1007/978-1-4612-4380-9_6.
[16] K.M. Pierce, J.C. Hoggard, J.L. Hope, P.M. Rainey, A.N. Hoofnagle, R.M. Jack, B.W. Wright, R.E. Synovec, Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts, Anal. Chem. 78 (2006) 5068–5075. https://doi.org/10.1021/ac0602625.
[17] LECO Corporation, ChromaTOF Tile Analytical Software, (n.d.). https://www.leco.com/product/chromatof-tile.
[18] N. Koljančić, A.A. Gomes, I. Špánik, A non-target geographical origin screening of botrytized wines through comprehensive two-dimensional gas chromatography coupled with high-resolution mass spectrometry, J. Sep. Sci. 46 (2023) 2300249. https://doi.org/10.1002/jssc.202300249.
[19] Y.V. Grishina, A.A. Vatlin, D.A. Mavletova, M.V. Odorskaya, A.M. Senkovenko, R.A. Ilyasov, V.N. Danilenko, Metabolites potentially determine the high antioxidant properties of limosilactobacillus fermentum U-21, BioTech 12 (2023) 39. https://doi.org/10.3390/biotech12020039.
[20] S.E. Prebihalo, G.S. Ochoa, K.L. Berrier, K.J. Skogerboe, K.L. Cameron, J.R. Trump, S.J. Svoboda, J.K. Wickiser, R.E. Synovec, Control-normalized Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data for enhanced biomarker discovery in a metabolomic study of orthopedic knee-ligament

injury, Anal. Chem. 92 (2020) 15526–15533. https://doi.org/10.1021/acs.analchem.0c03456.
[21] S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A 1667 (2022) 462868. https://doi.org/10.1016/j.chroma.2022.462868.
[22] M. Beccaria, C. Bobak, B. Maitshotlo, T.R. Mellors, G. Purcaro, F.A. Franchina, C.A. Rees, M. Nasir, A. Black, J.E. Hill, Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques, J. Breath Res. 13 (2018) 016005. https://doi.org/10.1088/1752-7163/aae80e.
[23] E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics, J. Chromatogr. A 1217 (2010) 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069.
[24] T. Sawoszczuk, J. Syguła-Cholewińska, J.M. del Hoyo-Meléndez, Optimization of headspace solid phase microextraction for the analysis of microbial volatile organic compounds emitted by fungi: Application to historical objects, J. Chromatogr. A 1409 (2015) 30–45. https://doi.org/10.1016/j.chroma.2015.07.059.
[25] C. Song, Y. Zhang, Q. Zhao, M. Chen, Y. Zhang, C. Gao, Z. Jia, S. Song, J. Guan, Z. Shang, Volatile organic compounds produced by Bacillus aryabhattai AYG1023 against Penicillium expansum causing blue mold on the Huangguan pear, Microbiol. Res. 278 (2024) 127531. https://doi.org/10.1016/j.micres.2023.127531.
[26] A. Marseglia, M. Musci, M. Rinaldi, G. Palla, A. Caligiani, Volatile fingerprint of unroasted and roasted cocoa beans (Theobroma cacao L.) from different geographical origins, Food Res. Int. 132 (2020) 109101. https://doi.org/10.1016/j.foodres.2020.109101.
[27] A. Hanifah, H. Firmanto, S.P. Putri, E. Fukusaki, Unique metabolite profiles of Indonesian cocoa beans from different origins and their correlation with temperature, J. Biosci. Bioeng. 134 (2022) 125–132. https://doi.org/10.1016/j.jbiosc.2022.05.001.
[28] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, R. McCrindle, M. Atlabachew, GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for tracing geographical origins from Ethiopia, J. Sci. Food Agr. 99 (2019) 3811–3823. https://doi.org/10.1002/jsfa.9603.
[29] L. Jirovetz, D. Smith, G. Buchbauer, Aroma compound analysis of Eruca sativa (Brassicaceae) SPME headspace leaf samples using GC, GC−MS, and olfactometry, J. Agric. Food Chem. 50 (2002) 4643–4646. https://doi.org/10.1021/jf020129n.
[30] A.M. Api, D. Belsito, D. Botelho, M. Bruze, G.A. Burton, J. Buschmann, M.L. Dagli, M. Date, W. Dekant, C. Deodhar, M. Francis, A.D. Fryer, L. Jones, K. Joshi, S. La Cava, A. Lapczynski, D.C. Liebler, D. O'Brien, A. Patel, T.M. Penning, G. Ritacco, J. Romine, N. Sadekar, D. Salvito, T.W. Schultz, I.G. Sipes, G. Sullivan, Y. Thakkar, Y. Tokura, S. Tsang, RIFM fragrance ingredient safety assessment, butyl benzoate, CAS registry number 136-60-7, Food Chem. Toxicol. 122 (2018) S680–S686. https://doi.org/10.1016/j.fct.2018.11.025.
[31] X. Shi, S. Zhao, S. Chen, X. Han, Q. Yang, L. Zhang, X. Xia, J. Tu, Y. Hu, Tetramethylpyrazine in Chinese baijiu: presence, analysis, formation, and regulation, Front.

Nutr. 9 (2022). https://www.frontiersin.org/articles/10.3389/fnut.2022.1004435 (accessed December 22, 2023).
[32] F. Mohamadi Alasti, N. Asefi, R. Maleki, S.S. SeiiedlouHeris, Investigating the flavor compounds in the cocoa powder production process, Food Sci. Nutr. 7 (2019) 3892–3901. https://doi.org/10.1002/fsn3.1244.
[33] E.E. Yost, S.Y. Euling, J.A. Weaver, B.E.J. Beverly, N. Keshava, A. Mudipalli, X. Arzuaga, T. Blessinger, L. Dishaw, A. Hotchkiss, S.L. Makris, Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies, Environ. Int. 125 (2019) 579–594. https://doi.org/10.1016/j.envint.2018.09.038.
[34] Y. Yan, F. Zhu, C. Zhu, Z. Chen, S. Liu, C. Wang, C. Gu, Dibutyl phthalate release from polyvinyl chloride microplastics: Influence of plastic properties and environmental factors, Water Res. 204 (2021) 117597. https://doi.org/10.1016/j.watres.2021.117597.
[35] E.T. Contis, C.-T. Ho, C.J. Mussinan, T.H. Parliment, F. Shahidi, A.M. Spanier, Food Flavors: Formation, Analysis and Packaging Influences, Elsevier, 1998.
[36] N. Tao, Y. Chen, Y. Wu, X. Wang, L. Li, A. Zhu, The terpene limonene induced the green mold of citrus fruit through regulation of reactive oxygen species (ROS) homeostasis in Penicillium digitatum spores, Food Chem. 277 (2019) 414–422. https://doi.org/10.1016/j.foodchem.2018.10.142.
[37] M. Jordán, K. Goodner, P. Shaw, Volatile components in banana (Musa acuminata colla cv. Cavendish) and yellow passion fruit (Passiflora edulis Sims / flavicarpa Degner) as determined by GC-MS and GC-olfactometry, Proc. Fla. State Hort. Soc. 114 (2001) 153-157
[38] E. Sheibani, S.E. Duncan, D.D. Kuhn, A.M. Dietrich, J.J. Newkirk, S.F. O'Keefe, Changes in flavor volatile composition of oolong tea after panning during tea processing, Food Sci. Nutr. 4 (2016) 456–468. https://doi.org/10.1002/fsn3.307.
[39] S.H. Hazeena, R. Sindhu, A. Pandey, P. Binod, Lignocellulosic bio-refinery approach for microbial 2,3-butanediol production, Bioresour. Technol. 302 (2020) 122873. https://doi.org/10.1016/j.biortech.2020.122873.
[40] J. Rodriguez-Campos, H.B. Escalona-Buendía, I. Orozco-Avila, E. Lugo-Cervantes, M.E. Jaramillo-Flores, Dynamics of volatile and non-volatile compounds in cocoa (Theobroma cacao L.) during fermentation and drying processes using principal components analysis, Food Res. Int. 44 (2011) 250–258. https://doi.org/10.1016/j.foodres.2010.10.028.
[41] L.G. Dias, A. Hacke, S.F. Bergara, O.V. Villela, L.R.B. Mariutti, N. Bragagnolo, Identification of volatiles and odor-active compounds of aromatic rice by OSME analysis and SPME/GC-MS, Food Res. Int. 142 (2021) 110206. https://doi.org/10.1016/j.foodres.2021.110206.
[42] R.J. Peterson, S.S. Chang, Identification of volatile flavor compounds of fresh, frozen beef stew and a comparison of these with those of canned beef stew, J. Food Sci. 47 (1982) 1444–1448. https://doi.org/10.1111/j.1365-2621.1982.tb04957.x.
[43] C. Chen, W. Zhou, H. Yu, J. Yuan, H. Tian, Evaluation of the perceptual interactions among aldehydes in a cheddar cheese matrix according to odor threshold and aroma intensity, Molecules 25 (2020) 4308. https://doi.org/10.3390/molecules25184308.
[44] M.S. Gill, A.J. Macleod, M. Moreau, Volatile components of cocoa with particular reference to glucosinolate products, Phytochemistry 23 (1984) 1937–1942. https://doi.org/10.1016/S0031-9422(00)84945-6.

[45] M.K. Syrokou, S. Paramithiotis, C.D. Kanakis, G.K. Papadopoulos, P.A. Tarantilis, P.N. Skandamis, L. Bosnea, M. Mataragas, E.H. Drosinos, Effect of dough-related parameters on the antimold activity of Wickerhamomyces anomalus strains and mold-free shelf life of bread, Appl. Sci. 12 (2022) 4506. https://doi.org/10.3390/app12094506.

Chapter 7. Tile-based analysis with support vector machine regression modeling of GC×GC-TOFMS data of VOCs produced by *Malassezia pachydermatis* grown at variable pHs

7.1 Introduction

Malassezia pachydermatis is a species of yeast that can be found on mammalian skin. *Malassezia* species are lipid dependent and inhabit sebaceous areas, such as the scalp, face, or behind the ears, at which these yeasts can intake lipids from their host [1,2]. *Malassezia* is a ubiquitous commensal microbe that can also be an opportunistic yeast where under certain microenvironmental conditions the establishment of dermatologic and systemic diseases in the host can be triggered [2–7].

Volatile organic compounds (VOCs) are carbon-based compounds with a low molecular weight and, generally, a high vapor pressure [6]. Several factors are involved in specifically producing microbial VOCs, including substrate and temperature. Further, diversity in the composition of VOCs can be altered by factors such as pH [3,4]. Evidence suggests microbial VOCs are generated from central and secondary metabolisms, therefore, chemical communication mediated by VOCs released by microorganisms is pivotal to understanding what triggers the yeast to become pathogenic. Currently, there are studies starting to understand the metabolic pathways of VOCs of *Malassezia* on a genus level; however, untangling the complex volatilome and the pathogenic triggers of *Malassezia*, and microbes in general, is challenging.

For the chemical analysis of such complex systems, comprehensive two-dimensional (2D) gas chromatography with time-of-flight mass spectrometry (GC×GC-TOFMS) has been growing in use in various areas including screening metabolite changes in specific diseases like tuberculosis [8,9], and SARS-CoV-2 [10,11] as well as changes in breath based on changes in

diet [12]. GC×GC-TOFMS allows for a nontargeted and high-throughput analysis of samples, which is incredibly important when analyzing biological samples, where there are many analytes present and interacting with one another in various ways [13,14]. Due to the complex nature of GC×GC-TOFMS datasets, it is hard to interpret them manually. Usually, advanced software "chemometric" methods are used for extracting chemical information from such datasets, with the least possible interaction from the user [15].

There are two general categories of chemometric methods often implemented: targeted and untargeted, with the latter focusing on all of the analytes that are responsible for differences or similarities between classes of samples. Further, these methods can be supervised or unsupervised, where supervised methods rely on knowledge of sample classes before analysis. A powerful supervised untargeted method is tile-based Fisher ratio (F-ratio) analysis [16,17] used here to find class distinguishing analytes between *M. pachydermatis* grown in three pH growth media: 5.7, 9.7, and 12.4. The F-ratio is calculated as the ratio of between-class variance relative to the sum of within-class variance within a small 2D portion of the chromatogram, called a tile, on a mass channel (*m/z*) basis [16,17]. After F-ratio analysis, the user is provided with a hit list, with analyte hits ranked in descending order of their F-ratio. The higher the F-ratio, the more likely the analyte hit is going to be statistically significantly different in concentration between the two or more sample classes, so the most informative analytes should be near the top of the hit list. After F-ratio analysis, the user can readily extract the signal for every analyte for each sample class. By using analyte signals, further visualization and evaluation of relationships between classes can be performed, using methods like PCA [18–21] or k-means [22,23]. However, in some cases these methods do not provide clear relationships between the sample

classes. To address this challange, post-processing of the analyte signals may need to be performed, e.g., either on simulated data [13] or by applying suitable quantitative metrics [24].

Herein, we present a study of VOCs produced by *M. pachydermatis* grown at three pHs that is initially confounded by interwoven sample class relationships. Due to the changing VOC profiles produced by media blanks when their pH is changed, they were also analyzed. Because of that, this dataset had 6 sample classes, and for each class, 5 replicates were collected. Overall, 30 GC×GC-TOFMS chromatograms were analyzed using tile-based F-ratio analysis, after which the signal pattern for each analyte across the 6 sample classes, deemed significant by passing a suitable F-ratio threshold based on previous work [17]**,** were extracted and analyzed for further classification using two quantitative metrics (*R* and *RSD*). The *R* metric was defined to focus upon the differences between analyte signals of blanks and *M. pachydermatis* by taking away the influence of pH changes. The *RSD* metric, on the other hand, was defined to evaluate only the influence of pH. Based on the *R* metric magnitude, the analytes were split into 3 categories: media analytes consumed (in media only), pH-dependent analytes (in media and *M. pachydermatis*), and analytes produced in *M. pachydermatis* only. Using these metric assessments, we also built a support vector machine model that was used to confirm that the metrics we defined and implemented were a valid approach to catagorize analytes. The model will be demonstrated to provide high selectivity and specificity with a very low error. To our knowledge, this is the first study analyzing the VOC profile of *M. pachydermatis* using GC×GC-TOFMS, as well as looking into changes in the VOC profiles when pH of culture conditions is changing for *M. pachydermatis*.

7.2 Experimental

7.2.1 Malassezia pachydermatis microbial culture

The strain CBS 1879 of *M. pachydermatis* was purchased from the American Type Culture Collection (ATCC, Manassas, Virginia) for this study. *M. pachydermatis* was maintained in modified Dixon agar (mDixon per liter: 20 g noble agar, 36 g malt extract, 10 g mycological peptones (oxoid L40), 10 g desiccated oxbile, 10 mL Tween 60, 4 mL glycerol 50%) and incubated for 48 hr at 31 ºC. Liquid cultures were prepared by removing an individual fungal colony from the agar plate and mixing into 10 mL of the liquid mDixon media in an autoclaved Erlenmeyer flask. The flask was then incubated on an orbital shaker (180 rpm) at 31 ºC for 48 hr.

7.2.2 Device Inoculation and Incubation

Prior to inoculation, 20 mL glass GC-MS vials (Supelco, St. Louis, MO) were ultrasonicated with 70% ethanol for 30 min and autoclaved at 121 ºC for 30 min on a gravity cycle. Metal caps with a rubber septum (Supelco, St. Louis, MO) for the vials were also sonicated in 70% ethanol for 30 min and dried with compressed air. The vials had 1 mL of mDixon agar (5.7 pH) adjusted accordingly to 9.7 and 12.4 pH with 1 N sodium hydroxide (NaOH) per experimental design for inoculation and blanks. After the 48 hr incubation period, the liquid fungal cultures were adjusted to 1×10^6 colony forming units per milliliter (CFU/mL) by optical density (OD600) analysis before inoculation onto the agar of the GC-MS vials. Each vial was inoculated with 10 µL of the liquid culture. The vials were capped and para filmed. Vials were then placed in a pipette tip box and placed in a 31 ºC incubator for an incubation period of 72 hr to allow fungal colony growth before sampling by headspace (HS) solid phase microextraction (SPME) followed by GC×GC-TOFMS data collection.

7.2.3 HS-SPME-GC×GC-TOFMS instrument conditions

Data was collected using a Pegasus BT 4D GC×GC-TOFMS with a cryogenic modulator (LECO, St. Joseph, MI). The headspace of each blank and Malassezia sample was sampled using a divinylbenzene/carboxen/polydimethylsiloxane SPME fiber (DVB/CAR/PDMS, fiber thickness: 50/30 μm). This fiber was chosen due to its ability to extract analytes of a wide boiling point and compound polarity range that were anticipated in these samples.

For the initial conditioning of a new DVB/CAR/PDMS fiber, it was held in a 250 °C GC inlet for 1 hr. Before each sample extraction, Malassezia samples and blanks were incubated at 31 °C for 5 min. The headspace of the samples was extracted for 30 min at 31 °C. This temperature was chosen to not alter *M. pachydermatis* samples by accidentally killing them at too high of a temperature. The SPME fiber with the extracted volatiles was desorbed splitless in the GC inlet for 5 min at 250 °C. Between sample extractions and chromatographic runs, the fiber was re-conditioned at 250 °C for 5 min. The sample extraction method utilized the L-PAL3 autosampler (LECO, St. Joseph, MI, USA).

Separations of the 6 samples in true quintuple (each replicate was made separately) were collected using the LECO Pegasus BT 4D GC×GC-TOFMS equipped with an Agilent 7890 GC (Agilent Technologies, Palo Alto, CA, USA) and a stock quad-jet thermal modulator. Splitless sample injections were separated on a non-polar Rxi-5Sil MS ^{1}D column (60 m × 0.25 mm × 0.25 μm; Restek), and a mid-polar Rtx-200 ^{2}D column (3 m × 0.18 mm × 0.2 μm; Restek). The ^{1}D column was held at 40 °C for 5 min before ramping to 140 °C at 8 °C/min, and then ramping to 250 °C at 30 °C/min where it was held for 10 min. The same temperature program was used for the ^{2}D oven and modulator with an offset of +5 °C and +15 °C, respectively. The carrier gas, ultra-high purity helium (Grade 5, 99.999 %, Praxair, Seattle, WA, USA), operated at a constant

flow rate of 2 mL/min. The ^{1}D effluent was reinjected on the ^{2}D column at a modulation period of 3 s. The ion source and transfer line temperatures were set to 225 °C and 285 °C, respectively. The TOFMS collected *m/z* 45–334 at 100 Hz with an electron ionization energy of 70 eV after a 10 s acquisition delay.

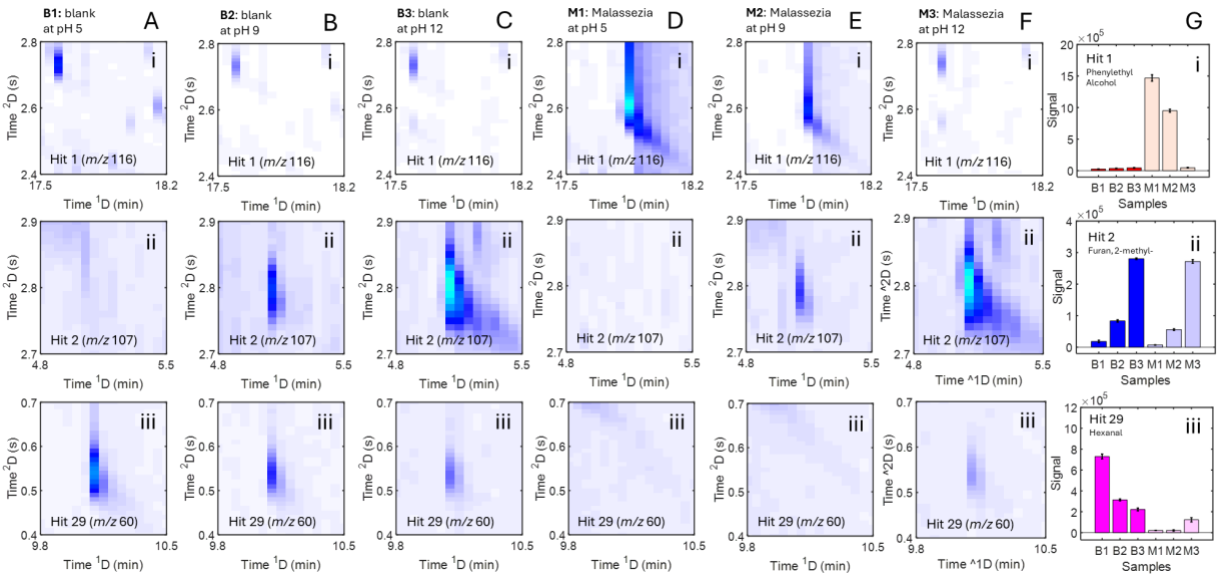


Figure 7.1. Zoom-in of selective ion chromatograms for (A) blank at pH 5, (B) blank at pH 9, (C) blank at pH 12, (D) Malassezia at pH 5, (E) Malassezia at pH 9, (F) Malassezia at pH 12 of (i) hit 1 at m/z 116, (ii) hit 2 at m/z 107 and (iii) hit 29 at m/z 60. (G) summed signal profile for (i) hit 1, (ii) hit 2 and (iii) hit 29. Error bars in G i-iii represent standard error of the mean of five replicates.

7.2.4 Data analysis.

GC×GC-TOFMS chromatograms were imported into ChromaTOF Tile Software (v. 101, LECO, St. Joseph, MI, USA) to perform tile-based Fisher ratio analysis, using ^{1}D tile size of 5 modulations (15 s) and ^{2}D tile size of 15 spectra (15 ms), where the hit list is ranked using the *m/z* signal that produced the highest F-ratio for each hit. The signal-to-noise (*S/N*) threshold was 10 and minimum masses per tile were set to 3. All the hits with an F-ratio ≥ 20 based on previous work [17], were investigated, and after artifact removal 566 analyte hits remained. Signals for each of those hits at their highest F-ratio *m/z* for the 30 chromatograms (3 pHs × 2 conditions (blank vs. *M. pachydermatis*) × 5 replicates) were transferred out of ChromaTOF Tile Software

and all the further data analysis was performed using Matlab 2023a using the Parallel Computing Toolbox (Mathworks, Inc., Natick, MA, USA). Support Vector Machine (SVM) Regression modeling was performed using PLS Toolbox (Eigenvector Research, Inc., Manson, WA, USA). For the SVM regression modeling the averaged signals over 5 replicates for 566 hits were used. The model was trained on the signals of those 566 hits that were normalized to the highest value out of 6 chromatograms for each hit. For the model, radial basis function kernel was used, no compression was performed and standard parameters (Table C.1.) for gamma, cost, epsilon, and nu were used. Venetian blinds were used as a cross-validation method using 10 data splits.

7.3 Results and Discussion

GC×GC-TOFMS chromatograms of all the samples were collected (the TIC of one replicate of each of these is shown in Figures C.1-2.) and then processed using tile-based F-ratio software (described in the Experimental section). Figure 1 shows how zoom-in chromatograms appear for a media blank at pH 5.7, 9.7, and 12.4 (Figure 7.1. A-C, respectively), as well as *M. pachydermatis* at pH 5.7, 9.7, and 12.4 (Figure 7.1. D-F, respectively) for three analyte hits (i-iii). Tile-based F-ratio analysis revealed 566 analyte hits exhibited sufficient class-to-class signal differences (F-ratio ≥ 20) across the 6 sample classes. Following the F-ratio analysis, each analyte hit was reduced to its signal pattern, with examples provided in Figure 1G, where "B" represents media samples and "M" represents *M. pachydermatis* samples, and then "1" represents pH 5.7, "2" represents pH 9.7 and "3" represents pH 12.4. Examples of the three distinctive signal patterns observed in this dataset are revealed in Figure 7.1. Gi-iii. For example, the analyte in Figure 7.1. Gi of phenylethyl alcohol (Hit 1) has a very small signal in all the blanks and a much higher signal in at least one of the *M. pachydermatis* samples. This is an example of an analyte produced by the *M. pachydermatis*, but with pH dependence. A different

signal pattern is revealed in Figure 7.1. Gii, which is 2-methylfuran (Hit 2), whereby the signals change only with pH, but do not change between blank and *M. pachydermatis*. This would be an example of the pH-only dependent analyte. Finally, there is a third signal pattern observed in Figure 7.1. Giii, which is hexanal (Hit 3), that exhibits high signal(s) in the media blanks, and low signal(s) in *M. pachydermatis* samples. This indicates an analyte that is consumed by *M. pachydermatis*.

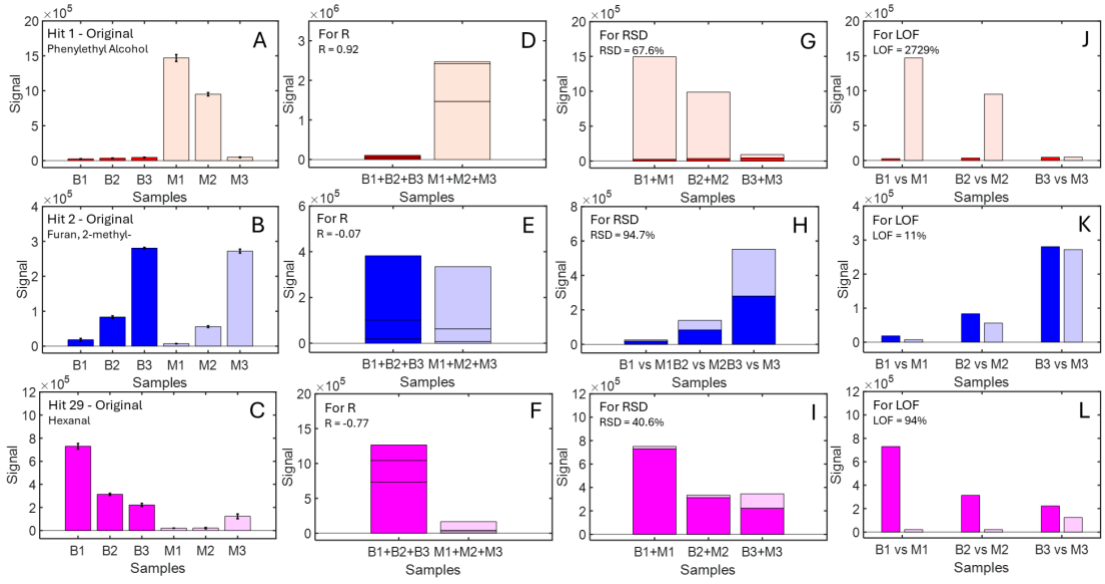


Figure 7.2. Original signal profiles for (A) hit 1, (B) hit 2, and (C) hit 29. Explanation of calculations of (D-F) R value, (G-I) RSD, and (J-L) LOF. Error bars in A-C represent standard error of the mean of five replicates.

As there were visually recognizable distinctive types of analyte signal patterns (Figure 7.2. A-C), we created a mathematical method to categorize them into unique groups for the purpose of uncoupling the effect of pH influence vs. media and *M. pachydermatis*. The first quantitative metric, *R*, was defined so as to remove the pH influence while focusing on the signal differences between blanks vs. *M. pachydermatis*. This *R* metric (Figure 7.2. D-F) is the ratio of sum of the blank signals across every pH minus the sum of the *M. pachydermatis* signals across every pH, normalized by the sum of these two summed quantities:

$$R\ metric = \frac{\sum_{i=1}^{3} M_i - \sum_{i=1}^{3} B_i}{\sum_{i=1}^{3} M_i + \sum_{i=1}^{3} B_i}$$

where M represents all *M. pachydermatis* signals and B represents all the media blank signals. As defined, the *R* metric can range from -1 to 1. The second quantitative metric, *RSD metric* (Figure 7.2. G-I), was defined so as to focus on the pH influence while minimizing the influence of the signal differences between blanks vs. *M. pachydermatis*, by taking a relative standard deviation (RSD) of the sum of blank and *M. pachydermatis* at every pH (B1+M1, B2+M2, B3+M3). The general equation for this *RSD* metric is shown below, where *STD* stands for standard deviation:

$$RSD = \frac{STD\ of\ B_1 + M_1, B_2 + M_2, \text{and } B_3 + M_3}{mean\ of\ B_1 + M_1, B_2 + M_2, \text{and } B_3 + M_3}$$

Lastly, we also utilized a lack-of-fit (*LOF*) metric (Figure 2J-L) using the following formula:

$$LOF\ = 100\% \sqrt{\frac{\sum_{i=1}^{3} M_i - B_i}{\sum_{i=1}^{3} B_i^2}}$$

The *LOF* metric determines to what extent the signal pattern for the media matches with the signal pattern for the *M. pachydermatis* across the pHs. Thus, for a given analyte the *LOF* collectively compares the blank to *M. pachydermatis* at pH 5.7, the blank to *M. pachydermatis* at pH 9.7, and the blank to the *M. pachydermatis* at pH 12.4. While we did not use *LOF* as a determinant of analyte signal pattern categorization, it was used to assist visualization and interpretation of the other two metrics: *R* and *RSD*. The *R*, *RSD*, and *LOF* for each of the 566 analyte hits were plotted (Figure 7.3. A-C) in order to visualize the categorization process for this dataset, principally relying upon the *R* metric which minimizes the influence of pH. We observe that a high positive *R* indicates the analyte is produced by *M. pachydermatis* (Figure 7.2. A,D for Hit 1, R = 0.92). Conversely, a high negative *R* indicates the analyte is consumed by *M.*

pachydermatis (Figure 7.2. C,F for Hit 29, R = -0.77). Finally, the third category has a small *R* (± surrounding 0). These analytes are only dependent on pH and do not change between the media blank and *M. pachydermatis* (Figure 7.2. B,E for Hit 2, R = -0.07). Next, we initially empirically examined the concept of applying *R* metric thresholds to separate these 3 categories. This initial *R* threshold categorization approach was more rigorously examined using a support vector machine (SVM) modeling approach, *vide infra*. For this initial categorization, for analytes that exhibit signal patterns that do not change significantly between the media blank and *M. pachydermatis*, with small *R* (± surrounding 0), we empirically define this condition as when analytes exhibit less than a 2-fold difference between the sum of the blank signals vs. the sum of the *M. pachydermatis* signals (meaning M=2B):

$$R\ metric = \frac{2B - B}{2B + B} = \frac{B}{3B} = 0.33$$

Accordingly, this condition results in a *R* metric threshold of R=±0.33. When this R threshold is applied (Figure 7.3. A-C), the 3 analyte signal pattern categories emerge: analytes produced by *M. pachydermatis* are in red, analytes consumed by *M. pachydermatis* are in pink, and analytes little to no change between blank and *M. pachydermatis* are in blue. For comparison, we also applied PCA to attempt to visualize the classification of the analyte signal pattern categories (Figure 7.3. D). While there is some classification clustering that appears to be emerging in Figure 7.3. D, the signal pattern categorization was not well suited to PCA, but instead was much better suited to the method presented herein using the *R* metric, with support from the *RSD* and *LOF* metrics. We believe the success of these quantitative metrics can be attributed to their underlying mathematical definition and how these definitions focus upon teasing apart meaningful information in tandem with the experimental design with 6 inter-related sample classes.

While the *R* metric is the key categorization metric to classify the analytes according to their trend between blanks and *M. pachydermatis* samples, the *RSD* metric describes how large of the variation in signal with pH changes in the samples (Figure 7.4.). Examples of analyte signal patterns with high *RSD* are on the top of the graph (1-methythiopropane, 2,3-dimethyl-2-cyclopentene-1-one, and 2-methyl-1-butanol) show a high influence of pH on their signal patterns. The low *RSD* examples on the bottom of the graph (octanal, 2,5-dimethylpyrazine, and 3-methyl-2-pentanone) show a relatively lower impact of pH on their signal patterns.

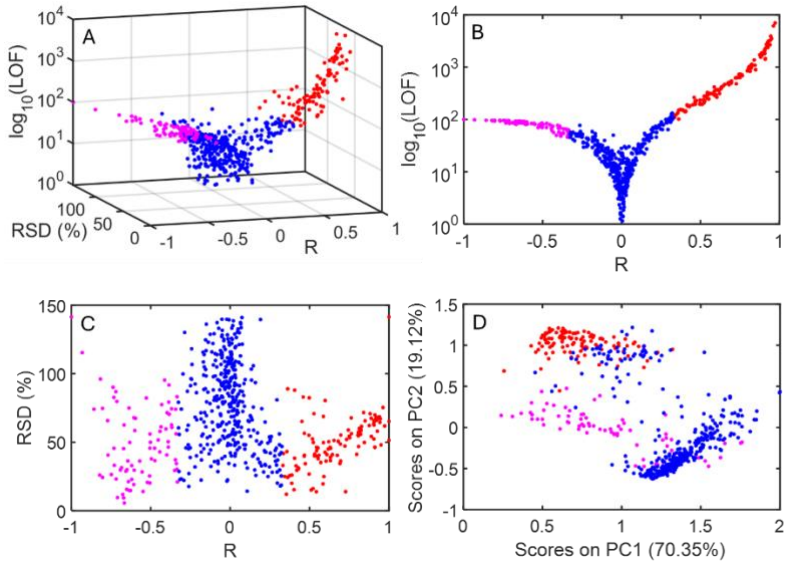


Figure 7.3. (A) Relationship between R, RSD, and the log of LOF for 566 analytes found through tile-based F-ratio analysis. (B) Relationship of R and LOF for 566 analytes. (C) Relationship of R and RSD for 566 analytes. (D) PCA plot made using signal profiles of 566 analytes. Color coding represents the class that the analyte was assigned based on its R value, where pink are analytes being consumed, blue are pH dependent analytes and red are analytes that are being produced by Malassezia.

The two examples of analytes dependent on pH (Hit 218 and 229) as well as the other two that are produced (Hit 450 and 501) have been reported in the literature in other species of *Malassezia* (*M. globosa*, *M. restricta*, *M. sympodialis*) than the one investigated here [3]. Additionally, 2,5-dimethylpyrazine (Hit 229) is one of the analytes that changes slightly with pH, and as we can see in Figure 7.4., it is present in both media and *M. pachydermatis* samples, which makes sense as in the literature it has been noted to exist in all eukaryotes, including

plants and animals, and the media components are derived from animal sources [25]. 2,5-dimethylpyrazine has also been found as a fermentation byproduct after inoculation with a different yeast of *Saccharomyces cerevisiae* [26]. The other analyte example of a pH dependent compound is 2,3-dimethyl-2-cyclopenten-1-one (Hit 218), which has been found as a natural product in *Mangifera indica* commonly called mango [27]. One of the examples of the analytes that get produced by *M. pachydermatis* is 2-methyl-1-butanol (Hit 450), which is a compound that is produced by many different microbes including *S. cerevisiae* [28], *Corynebacterium glutamicum* [29], *Escherichia coli* [30], and others [31]. It also is one of the compounds involved in signaling in microbe-to-microbe interactions [6]. 2-methyl-1-butanol is a promising alternative to ethanol as a biofuel, so production by a microbe could be very efficient [28]. The other example is 3-methyl-2-pentanone (Hit 501), which is produced by bacterial strains of *Microbacterium foliorum* [32], and *Pseudomonas aeruginosa* [33]. It has also been found in volatile space of both unspoiled and spoiled Iberian dry-cured hams [34]. Lastly, the two analyte examples of consumed analytes are octanal (Hit 405) and 1-methylthiopropane (Hit 80). Octanal has been previously find present in different growth mediums and shown to be consumed by *E. coli* [35].

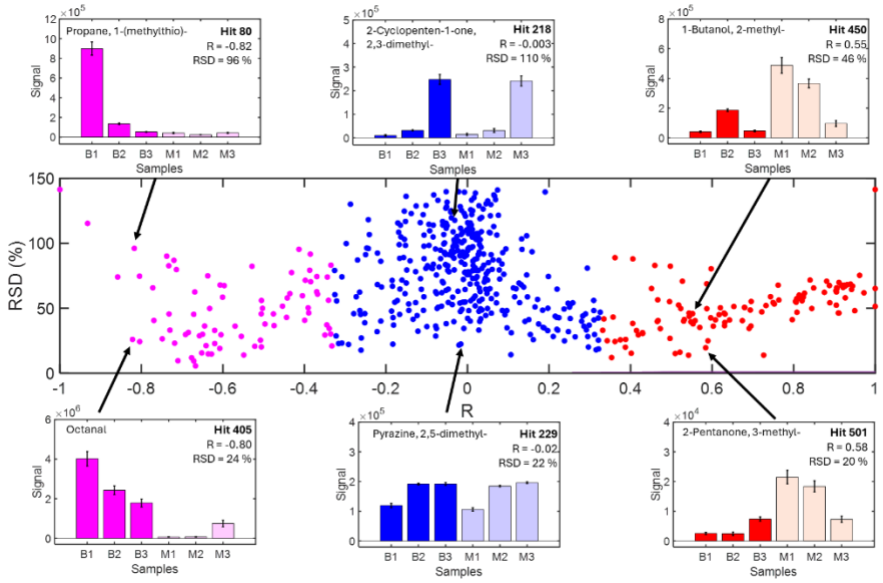


Figure 7.4. Plot of R vs RSD for 566 analytes, showing signal profile examples for analytes at similar R values, but different RSD values for three different classes of analytes. Error bars in signal patterns represent standard error of the mean of five replicates.

The top 20 identified analytes from each analyte category are shown in Figure 7.5. A full list of analytes is shown in Table C.2. The analytes that are consumed by *M. pachydermatis* are shown in Figure 7.5. A on a heatmap, where red represents high signal intensity, and blue represents low signal. We can see that most analytes are highest in the media of pH 5.7, however, there are a few analytes that are highest in the media of pH 9.7. The analytes that are pH dependent are shown in Figure 7.5. B and the analytes that are produced are shown in Figure 7.5. C. From the analytes produced many have been found previously, including phenylethyl alcohol, 1-pentanol, 3-methyl-1-butanol, and 2-ethyl-1-hexanol [3]. Another analyte, ethanol, has been found to be produced by the fungus *Muscodor albus,* which is known to release volatiles that inhibit other fungi, bacteria, and insects [36]. There is also evidence that microbes can convert fatty acids (that are present in the media) to corresponding alcohols [37], like 1-pentanol (Hit 88 – Figure 7.5. C), or benzyl alcohol (Hit 160 – Figure 7.5. C).

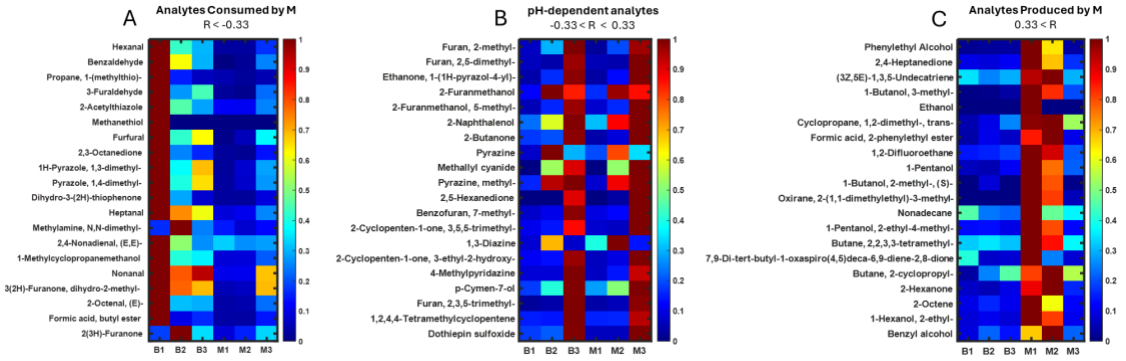


Figure 7.5. Signal profiles of top 20 identified analytes of each class: (A) for consumed analytes, (B) for analytes that are pH dependent and (C) for produced analytes.

Lastly, we verified that our way of assigning analytes into three groups was valid, by creating a support vector machine (SVM) model. We took all the 566 analyte signal patterns (normalized to the maximum value of each to be able to compare them to one another fairly) with their respective assigned class of consumed, pH dependent or produced and submitted that to a SVM model. Using venetian blinds, we created a cross-validation (CV) set. Results of CV are shown in Table 7.1. Three metrics we use for evaluation of the model are: true positive rate (TPR), true negative rate (TNR), and error. True positive rate (TPR), also called sensitivity is calculated using the following:

$$TPR = \frac{TP}{TP + FN}$$

Where TP is the number of true positives and FN is the number of false negatives. This is calculating the proportion of positive cases that were correctly identified. In our case all the three analyte catagories are performing very well with all of them having a TPR of 0.95 or higher, with 1 being perfect. True negative rate (TNR), also called specificity is calculated using the following:

$$TNR = \frac{TN}{TN + FP}$$

Where TN is the number of true negatives and FP is the number of false positives. This is calculating the proportion of negative cases that were classified correctly. Again, for our model, all three analyte categories have TNR of 0.95 or higher, with 1 being perfect. Lastly, error is calculated using:

$$Error = \frac{FP + FN}{TP + TN + FP + FN}$$

This is calculating the proportion of samples that were incorrectly classified. In our case, the highest error was 0.026, which is 2.6%, and is considerably very low with the other 2 groups being even lower around 1%. This result is very exciting, as it shows that our classification based on the *R* metric has a basis and a blind classification model classified the analytes into the same assignments.

Table 7.1. Summary of results of support vector machine regression model using cross-validation results.

Class	TPR	TNR	Error
Consumed Analytes	0.965	0.996	0.011
pH-Dependent Analytes	0.981	0.959	0.027
Produced Analytes	0.949	0.990	0.016

7.4 Conclusion

Here we present analysis of the volatile space of *M. pachydermatis* at the three pHs (5.7, 9.7, and 12.4) by GC×GC-TOFMS. For this analysis, we also analyzed media blanks at the three pHs, with samples and blanks collected in 5 replicates. Following the analysis, the chromatograms were analyzed using F-ratio software that found 566 analytes, out of which 288

had an identification following a NIST library search. For those 566 analytes signal patterns were calculated by averaging replicates, and two metrics were calculated: R and RSD. R was a ratio metric that only evaluated what was changing between blanks and *M. pachydermatis* by removing the influence of pH. RSD on the other hand, was used to evaluate the influence of pH. Based on R value, the analytes were split into 3 categories: media analytes consumed, pH dependent analytes, and *M. pachydermatis* produced analytes. Many of the produced analytes were already shown to be produced by other yeast species and shown to have biological significance. Further, there is evidence of some bioconversions between the analytes we found consumed and those we found produced. We also verified our classification results using support vector machines (SVM) model, where cross validation showed very promising results with TPR and TNR both being over 0.95 and error being below 0.03 (or 3 %). To our knowledge, this is the first study analyzing the *M. pachydermatis* volatome using GC×GC-TOFMS allowing for a more in depth look at the volatile signals being consumed as well as produced by the microbe. Microenvironmental changes including pH are hypothesized to trigger *M. pachydermatis* commensal to pathogenic transitions. It is important to analyze how the VOC profile changes with pH change to better characterize metabolic processes in *M. pachydermatis* in different microenvironments.

7.5 References

[1] A.M. Celis Ramírez, A. Amézquita, J.E.C. Cardona Jaramillo, L.F. Matiz-Cerón, J.S. Andrade-Martínez, S. Triana, M.J. Mantilla, S. Restrepo, A.F.G. Barrios, H. de Cock, Analysis of Malassezia Lipidome Disclosed Differences Among the Species and Reveals Presence of Unusual Yeast Lipids, Front Cell Infect Microbiol 10 (2020) 338. https://doi.org/10.3389/fcimb.2020.00338.

[2] A. Peano, E. Johnson, E. Chiavassa, P. Tizzani, J. Guillot, M. Pasquetti, Antifungal Resistance Regarding Malassezia pachydermatis: Where Are We Now?, J Fungi (Basel) 6 (2020) 93. https://doi.org/10.3390/jof6020093.

[3] A. Rios-Navarro, M. Gonzalez, C. Carazzone, A.M. Celis Ramírez, Why Do These Yeasts Smell So Good? Volatile Organic Compounds (VOCs) Produced by Malassezia Species in

the Exponential and Stationary Growth Phases, Molecules 28 (2023) 2620. https://doi.org/10.3390/molecules28062620.
[4] M. Gonzalez, A.M. Celis, M.I. Guevara-Suarez, J. Molina, C. Carazzone, Yeast Smell Like What They Eat: Analysis of Volatile Organic Compounds of Malassezia furfur in Growth Media Supplemented with Different Lipids, Molecules 24 (2019) 419. https://doi.org/10.3390/molecules24030419.
[5] A. Sastoque, S. Triana, K. Ehemann, L. Suarez, S. Restrepo, H. Wösten, H. de Cock, M. Fernández-Niño, A.F. González Barrios, A.M. Celis Ramírez, New Therapeutic Candidates for the Treatment of Malassezia pachydermatis -Associated Infections, Sci Rep 10 (2020) 4860. https://doi.org/10.1038/s41598-020-61729-1.
[6] A. Rios-Navarro, M. Gonzalez, C. Carazzone, A.M. Celis Ramírez, Learning about microbial language: possible interactions mediated by microbial volatile organic compounds (VOCs) and relevance to understanding Malassezia spp. metabolism, Metabolomics 17 (2021) 39. https://doi.org/10.1007/s11306-021-01786-3.
[7] J. Yang, S. Park, H.J. Kim, S.J. Lee, W.H. Jung, The Interkingdom Interaction with Staphylococcus Influences the Antifungal Susceptibility of the Cutaneous Fungus Malassezia, 33 (2023) 180–187. https://doi.org/10.4014/jmb.2210.10039.
[8] P.C.K. Makhubela, E.R. Rohwer, Y. Naudé, Detection of tuberculosis-associated compounds from human skin by GCxGC-TOFMS, Journal of Chromatography B 1231 (2023) 123937. https://doi.org/10.1016/j.jchromb.2023.123937.
[9] D. Beukes, M. van Reenen, D.T. Loots, I. du Preez, Tuberculosis is associated with sputum metabolome variations, irrespective of patient sex or HIV status: an untargeted GCxGC-TOFMS study, Metabolomics 19 (2023) 55. https://doi.org/10.1007/s11306-023-02017-7.
[10] A.Z. Berna, E.H. Akaho, R.M. Harris, M. Congdon, E. Korn, S. Neher, M. M'Farrej, J. Burns, A.R. Odom John, Reproducible Breath Metabolite Changes in Children with SARS-CoV-2 Infection, ACS Infect. Dis. 7 (2021) 2596–2603. https://doi.org/10.1021/acsinfecdis.1c00248.
[11] E. Barberis, E. Amede, S. Khoso, L. Castello, P.P. Sainaghi, M. Bellan, P.E. Balbo, G. Patti, D. Brustia, M. Giordano, R. Rolla, A. Chiocchetti, G. Romani, M. Manfredi, R. Vaschetto, Metabolomics Diagnosis of COVID-19 from Exhaled Breath Condensate, Metabolites 11 (2021) 847. https://doi.org/10.3390/metabo11120847.
[12] K. Raninen, R. Nenonen, E. Järvelä-Reijonen, K. Poutanen, H. Mykkänen, O. Raatikainen, Comprehensive Two-Dimensional Gas Chromatography–Mass Spectrometry Analysis of Exhaled Breath Compounds after Whole Grain Diets, Molecules 26 (2021) 2667. https://doi.org/10.3390/molecules26092667.
[13] L. Mikaliunaite, R.E. Synovec, Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, Talanta 244 (2022) 123396. https://doi.org/10.1016/j.talanta.2022.123396.
[14] L. Mikaliunaite, R.E. Synovec, Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data, Journal of Chromatography A 1730 (2024) 465093. https://doi.org/10.1016/j.chroma.2024.465093.
[15] T.J. Trinklein, C.N. Cain, G.S. Ochoa, S. Schöneich, L. Mikaliunaite, R.E. Synovec, Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis, Anal. Chem. 95 (2023) 264–286. https://doi.org/10.1021/acs.analchem.2c04235.

[16] L.C. Marney, W. Christopher Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895. https://doi.org/10.1016/j.talanta.2013.06.038.
[17] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819. https://doi.org/10.1021/ac504472s.
[18] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, Journal of Chromatography A 1662 (2022) 462735. https://doi.org/10.1016/j.chroma.2021.462735.
[19] E.M. Humston-Fulmer, D.E. Alonso, J.E. Binkley, Chapter Six - Improving cannabis differentiation by expanding coverage of the chemical profile with GCxGC-TOFMS, in: I. Ferrer, E.M. Thurman (Eds.), Comprehensive Analytical Chemistry, Elsevier, 2020: pp. 169–196. https://doi.org/10.1016/bs.coac.2020.04.007.
[20] K.A. Favela, M.J. Hartnett, J.A. Janssen, D.W. Vickers, A.J. Schaub, H.A. Spidle, K.S. Pickens, Nontargeted Analysis of Face Masks: Comparison of Manual Curation to Automated GCxGC Processing Tools, J. Am. Soc. Mass Spectrom. 32 (2021) 860–871. https://doi.org/10.1021/jasms.0c00318.
[21] C. Cheung, J.D. Baker, J.M. Byrne, K.A. Perrault, Investigating volatiles as the secondary metabolome of *Piper methysticum* from root powder and water extracts using comprehensive two-dimensional gas chromatography, Journal of Ethnopharmacology 294 (2022) 115346. https://doi.org/10.1016/j.jep.2022.115346.
[22] P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Analytica Chimica Acta 1209 (2022) 339847. https://doi.org/10.1016/j.aca.2022.339847.
[23] R. Costa, C. Fanali, G. Pennazza, L. Tedone, L. Dugo, M. Santonico, D. Sciarrone, F. Cacciola, L. Cucchiarini, M. Dachà, L. Mondello, Screening of volatile compounds composition of white truffle during storage by GCxGC-(FID/MS) and gas sensor array analyses, LWT - Food Science and Technology 60 (2015) 905–913. https://doi.org/10.1016/j.lwt.2014.09.054.
[24] C.N. Cain, T.J. Trinklein, G.S. Ochoa, R.E. Synovec, Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification, Anal. Chem. 94 (2022) 5658–5666. https://doi.org/10.1021/acs.analchem.2c00223.
[25] 2,5-Dimethylpyrazine (YMDB01598) - Yeast Metabolome Database, (n.d.). https://www.ymdb.ca/compounds/YMDB01598 (accessed July 18, 2024).
[26] R.F. Schwan, A.P.P. Bressani, S.J. Martinez, N.N. Batista, D.R. Dias, The essential role of spontaneous and starter yeasts in cocoa and coffee fermentation, FEMS Yeast Research 23 (2023) foad019. https://doi.org/10.1093/femsyr/foad019.
[27] PubChem, 2,3-Dimethyl-2-cyclopenten-1-one, (n.d.). https://pubchem.ncbi.nlm.nih.gov/compound/14270 (accessed July 18, 2024).
[28] Y. Zhang, S. Lane, J.-M. Chen, S.K. Hammer, J. Luttinger, L. Yang, Y.-S. Jin, J.L. Avalos, Xylose utilization stimulates mitochondrial production of isobutanol and 2-methyl-1-butanol in

Saccharomyces cerevisiae, Biotechnol Biofuels 12 (2019) 223. https://doi.org/10.1186/s13068-019-1560-2.
[29] M. Vogt, C. Brüsseler, J. van Ooyen, M. Bott, J. Marienhagen, Production of 2-methyl-1-butanol and 3-methyl-1-butanol in engineered *Corynebacterium glutamicum*, Metabolic Engineering 38 (2016) 436–445. https://doi.org/10.1016/j.ymben.2016.10.007.
[30] A.F. Cann, J.C. Liao, Production of 2-methyl-1-butanol in engineered Escherichia coli, Appl Microbiol Biotechnol 81 (2008) 89–98. https://doi.org/10.1007/s00253-008-1631-y.
[31] H. Su, H. Chen, J. Lin, Enriching the Production of 2-Methyl-1-Butanol in Fermentation Process Using Corynebacterium crenatum, Curr Microbiol 77 (2020) 1699–1706. https://doi.org/10.1007/s00284-020-01961-0.
[32] P. Deetae, P. Bonnarme, H.E. Spinnler, S. Helinck, Production of volatile aroma compounds by bacterial strains isolated from different surface-ripened French cheeses, Appl Microbiol Biotechnol 76 (2007) 1161–1171. https://doi.org/10.1007/s00253-007-1095-5.
[33] H.D. Bean, J.-M.D. Dimandja, J.E. Hill, Bacterial volatile discovery using solid phase microextraction and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry, Journal of Chromatography B 901 (2012) 41–46. https://doi.org/10.1016/j.jchromb.2012.05.038.
[34] A. Martín, M.J. Benito, E. Aranda, S. Ruiz-Moyano, J.J. Córdoba, M.G. Córdoba, Characterization by Volatile Compounds of Microbial Deep Spoilage in Iberian Dry-Cured Ham, Journal of Food Science 75 (2010) M360–M365. https://doi.org/10.1111/j.1750-3841.2010.01674.x.
[35] I.-A. Ratiu, T. Ligor, V. Bocos-Bintintan, H. Al-Suod, T. Kowalkowski, K. Rafińska, B. Buszewski, The effect of growth medium on an Escherichia coli pathway mirrored into GC/MS profiles, J. Breath Res. 11 (2017) 036012. https://doi.org/10.1088/1752-7163/aa7ba2.
[36] C.J. Alpha, M. Campos, C. Jacobs-Wagner, S.A. Strobel, Mycofumigation by the Volatile Organic Compound-Producing Fungus Muscodor albus Induces Bacterial Cell Death through DNA Damage, Appl Environ Microbiol 81 (2015) 1147–1156. https://doi.org/10.1128/AEM.03294-14.
[37] L.-J. Lin, M. Saini, C.-J. Chiang, Y.-P. Chao, Biocatalytic Conversion of Short-Chain Fatty Acids to Corresponding Alcohols in Escherichia coli, Processes 9 (2021) 973. https://doi.org/10.3390/pr9060973.

Chapter 8. Bibliography

2,5-Dimethylpyrazine (YMDB01598)—Yeast Metabolome Database. (n.d.). Retrieved July 18, 2024, from https://www.ymdb.ca/compounds/YMDB01598

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433–459. https://doi.org/10.1002/wics.101

Adahchour, M., Brandt, M., Baier, H.-U., Vreuls, R. J. J., Batenburg, A. M., & Brinkman, U. A. Th. (2005). Comprehensive two-dimensional gas chromatography coupled to a rapid-scanning quadrupole mass spectrometer: Principles and applications. *Journal of Chromatography A, 1067*(1), 245–254. https://doi.org/10.1016/j.chroma.2004.09.094

Adebo, O. A., Oyeyinka, S. A., Adebiyi, J. A., Feng, X., Wilkin, J. D., Kewuyemi, Y. O., Abrahams, A. M., & Tugizimana, F. (2021). Application of gas chromatography–mass spectrometry (GC-MS)-based metabolomics for the study of fermented cereal and legume foods: A review. *International Journal of Food Science & Technology*, *56*(4), 1514–1534. https://doi.org/10.1111/ijfs.14794

Almstetter, M. F., Appel, I. J., Gruber, M. A., Lottaz, C., Timischl, B., Spang, R., Dettmer, K., & Oefner, P. J. (2009). Integrative Normalization and Comparative Analysis for Metabolic Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography−Time-of-Flight Mass Spectrometry. *Analytical Chemistry*, *81*(14), 5731–5739. https://doi.org/10.1021/ac900528b

Alpha, C. J., Campos, M., Jacobs-Wagner, C., & Strobel, S. A. (2015). Mycofumigation by the Volatile Organic Compound-Producing Fungus Muscodor albus Induces Bacterial Cell Death through DNA Damage. *Applied and Environmental Microbiology*, *81*(3), 1147–1156. https://doi.org/10.1128/AEM.03294-14

Amaral, M. S. S., Nolvachai, Y., & Marriott, P. J. (2020). Comprehensive two-dimensional gas chromatography advances in technology and applications: Biennial update. *Anal. Chem.*, *92*, 85–104. https://doi.org/10.1021/acs.analchem.9b05412

Antoniou, C. V., Koukouraki, E. E., & Diamadopoulos, E. (2006). Determination of chlorinated volatile organic compounds in water and municipal wastewater using headspace–solid phase microextraction–gas chromatography. *Journal of Chromatography A*, *1132*(1), 310–314. https://doi.org/10.1016/j.chroma.2006.08.082

Api, A. M., Belsito, D., Botelho, D., Bruze, M., Burton, G. A., Buschmann, J., Dagli, M. L., Date, M., Dekant, W., Deodhar, C., Francis, M., Fryer, A. D., Jones, L., Joshi, K., La Cava, S., Lapczynski, A., Liebler, D. C., O'Brien, D., Patel, A., Penning, T. M., Ritacco, G., Romine, J., Sadekar, N., Salvito, D., Schultz, T. W., Sipes, I. G., Sullivan, G., Thakkar, Y., Tokura, Y., & Tsang, S. (2018). RIFM fragrance ingredient safety assessment, butyl benzoate, CAS registry number 136-60-7. *Food and Chemical Toxicology, 122*, S680–S686. https://doi.org/10.1016/j.fct.2018.11.025

Auersvald, M., Šiman, M., Vozka, P., & Straka, P. (2024). Quantitative determination of olefins in pyrolysis oils from waste plastics and tires using selective adsorption by Ag-SiO_2 followed by GC×GC-FID. *Talanta, 245*, 126792. https://doi.org/10.1016/j.talanta.2024.126792

Bahaghighat, H. D., Freye, C. E., & Synovec, R. E. (2019). Recent advances in modulator technology for comprehensive two-dimensional gas chromatography. *TrAC Trends in Analytical Chemistry, 113*, 379–391. https://doi.org/10.1016/j.trac.2018.04.016

Bai, L., Smuts, J., Schenk, J., Cochran, J., & Schug, K. A. (2018). Comparison of GC-VUV, GC-FID, and comprehensive two-dimensional GC–MS for the characterization of weathered and unweathered diesel fuels. *Fuel*, *214*, 521–527. https://doi.org/10.1016/j.fuel.2017.11.053

Bai, L., Smuts, J., Walsh, P., Fan, H., Hildenbrand, Z., Wong, D., Wetz, D., & Schug, K. A. (2015). Permanent gas analysis using gas chromatography with vacuum ultraviolet detection. *Journal of Chromatography A*, *1388*, 244–250. https://doi.org/10.1016/j.chroma.2015.02.007

Barberis, E., Amede, E., Khoso, S., Castello, L., Sainaghi, P. P., Bellan, M., Balbo, P. E., Patti, G., Brustia, D., Giordano, M., Rolla, R., Chiocchetti, A., Romani, G., Manfredi, M., & Vaschetto, R. (2021). Metabolomics Diagnosis of COVID-19 from Exhaled Breath Condensate. *Metabolites*, *11*(12), Article 12. https://doi.org/10.3390/metabo11120847

Baylis, S. A., Hall, K., & Jumeau, E. J. (1994). The analysis of the C1–C5 components of natural gas samples using gas chromatography-combustion-isotope ratio mass spectrometry. *Organic Geochemistry*, *21*(6), 777–785. https://doi.org/10.1016/0146-6380(94)90019-1

Bean, H. D., Dimandja, J.-M. D., & Hill, J. E. (2012). Bacterial volatile discovery using solid phase microextraction and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry. *Journal of Chromatography B*, *901*, 41–46. https://doi.org/10.1016/j.jchromb.2012.05.038

Bean, H. D., Hill, J. E., & Dimandja, J.-M. D. (2015). Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography–mass spectrometry data. *Journal of Chromatography A*, *1394*, 111–117. https://doi.org/10.1016/j.chroma.2015.03.001

Beccaria, M., Bobak, C., Maitshotlo, B., Mellors, T. R., Purcaro, G., Franchina, F. A., Rees, C. A., Nasir, M., Black, A., & Hill, J. E. (2018). Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *Journal of Breath Research, 13*(1), 016005. https://doi.org/10.1088/1752-7163/aae80e

Beckstrom, A. C., Humston, E. M., Snyder, L. R., Synovec, R. E., & Juul, S. E. (2011). Application of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry method to identify potential biomarkers of perinatal asphyxia in a non-human primate model. *Journal of Chromatography A, 1218*(8), 1899–1906. https://doi.org/10.1016/j.chroma.2011.01.086

Bell, D. S., Synovec, R., & Mikaliunaite, L. (2022). Application of Porous Layer Open Tubular Columns: Beyond Permanent Gases. *LCGC North America*, *40*(10), 476–483. https://www.chromatographyonline.com/view/application-of-porous-layer-open-tubular-columns-beyond-permanent-gases

Bennett, D. (1967). Analysis of gas mixtures by gas chromatography. *Journal of Chromatography A*, *26*, 482–484. https://doi.org/10.1016/S0021-9673(01)98907-1

Berna, A. Z., Akaho, E. H., Harris, R. M., Congdon, M., Korn, E., Neher, S., M'Farrej, M., Burns, J., & Odom John, A. R. (2021). Reproducible Breath Metabolite Changes in Children with SARS-CoV-2 Infection. *ACS Infectious Diseases*, *7*(9), 2596–2603. https://doi.org/10.1021/acsinfecdis.1c00248

Beukes, D., van Reenen, M., Loots, D. T., & du Preez, I. (2023). Tuberculosis is associated with sputum metabolome variations, irrespective of patient sex or HIV status: An untargeted GCxGC-TOFMS study. *Metabolomics*, *19*(6), 55. https://doi.org/10.1007/s11306-023-02017-7

Bhide, A., Ganguly, A., Parupudi, T., Ramasamy, M., Muthukumar, S., & Prasad, S. (2021). Next-Generation Continuous Metabolite Sensing toward Emerging Sensor Needs. *ACS Omega*, *6*(9), 6031–6040. https://doi.org/10.1021/acsomega.0c06209

Biedermann, M., & Grob, K. (2019). Advantages of comprehensive two-dimensional gas chromatography for comprehensive analysis of potential migrants from food contact materials. *Analytica Chimica Acta, 1057*, 11–17. https://doi.org/10.1016/j.aca.2018.10.046

Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale, M. H., Trethewey, R. N., Lange, B. M., Wurtele, E. S., & Sumner, L. W. (2004). Potential of metabolomics as a functional genomics tool. *Trends in Plant Science*, *9*(9), 418–425. https://doi.org/10.1016/j.tplants.2004.07.004

Bordajandi, L. R., Ramos, J. J., Sanz, J., González, M. J., & Ramos, L. (2008). Comprehensive two-dimensional gas chromatography in the screening of persistent organohalogenated pollutants in environmental samples. *Journal of Chromatography A*, *1186*(1), 312–324. https://doi.org/10.1016/j.chroma.2007.12.013

Boswell, H., Tarazona Carrillo, K., & Górecki, T. (2020). Evaluation of the performance of cryogen-free thermal modulation-based comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC×GC-TOFMS) for the qualitative analysis of a complex bitumen sample. *Separations, 7*(1), 13. https://doi.org/10.3390/separations7010013

Brasseur, C., Dekeirsschieter, J., Schotsmans, E. M. J., de Koning, S., Wilson, A. S., Haubruge, E., & Focant, J.-F. (2012). Comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry for the forensic study of cadaveric volatile organic compounds released in soil by buried decaying pig carcasses. *Journal of Chromatography A*, *1255*, 163–170. https://doi.org/10.1016/j.chroma.2012.03.048

Brito, L., Payan, F., Albrieux, F., Guillon, E., Martens, J. A., & Pirngruber, G. D. (2023). Hydrocracking of a long chain alkyl-cycloalkane: Role of porosity and metal-acid balance. *ChemCatChem, 15*(8), e202201286. https://doi.org/10.1002/cctc.202201286

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*(2), 149–171. https://doi.org/10.1016/S0169-7439(97)00032-4

Brokl, M., Bishop, L., Wright, C. G., Liu, C., McAdam, K., & Focant, J.-F. (2014). Multivariate analysis of mainstream tobacco smoke particulate phase by headspace solid-phase micro extraction coupled with comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry. *Journal of Chromatography A*, *1370*, 216–229. https://doi.org/10.1016/j.chroma.2014.10.057

Brown, A. S., Milton, M. J. T., Cowper, C. J., Squire, G. D., Bremser, W., & Branch, R. W. (2004). Analysis of natural gas by gas chromatography: Reduction of correlated uncertainties by normalisation. *Journal of Chromatography A*, *1040*(2), 215–225. https://doi.org/10.1016/j.chroma.2004.04.007

Brown, S. D., Blank, T. B., Sum, S. T., & Weyer, L. G. (1994). Chemometrics. *Analytical Chemistry, 66*(14), 315–359. https://doi.org/10.1021/ac00084a014

Bruckner, C. A., Prazen, B. J., & Synovec, R. E. (1998). Comprehensive Two-Dimensional High-Speed Gas Chromatography with Chemometric Analysis. *Analytical Chemistry*, *70*(14), 2796–2804. https://doi.org/10.1021/ac980164m

Bruno, T. J. (2009). Simple, Quantitative Headspace Analysis by Cryoadsorption on a Short Alumina PLOT Column*. *Journal of Chromatographic Science*, *47*(7), 569–574. https://doi.org/10.1093/chromsci/47.7.569

Buldakov, M. A., Korolkov, V. A., Matrosov, I. I., Petrov, D. V., Tikhomirov, A. A., & Korolev, B. V. (2013). Analyzing natural gas by spontaneous Raman scattering spectroscopy. *Journal of Optical Technology*, *80*(7), 426–430. https://doi.org/10.1364/JOT.80.000426

Burger, J. L., Lovestead, T. M., & Bruno, T. J. (2016). Composition of the C6+ Fraction of Natural Gas by Multiple Porous Layer Open Tubular Capillaries Maintained at Low Temperatures. *Energy & Fuels*, *30*(3), 2119–2126. https://doi.org/10.1021/acs.energyfuels.6b00043

Burnetti, A. J., Aydin, M., & Buchler, N. E. (2016). Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates. *Molecular Biology of the Cell*, *27*(1), 64–74. https://doi.org/10.1091/mbc.E15-07-0454

Byrne, J. M., Dubois, L. M., Baker, J. D., Focant, J.-F., & Perrault, K. A. (2020). A non-targeted data processing workflow for volatile organic compound data acquired using comprehensive two-dimensional gas chromatography with dual channel detection. *MethodsX, 7*, 101009. https://doi.org/10.1016/j.mex.2020.101009

Cai, H., & Stearns, S. D. (2018). A comprehensive two-dimensional gas chromatography valve modulation method using hold-release primary column flow for long secondary separation time with 100% transfer. *Journal of Chromatography A*, *1569*, 200–211. https://doi.org/10.1016/j.chroma.2018.07.064

Cain, C. N., Gaida, M., Stefanuto, P.-H., Focant, J.-F., Synovec, R. E., Jackels, S. C., & Skogerboe, K. J. (2024). Investigating sensory-classified roasted arabica coffee with GC × GC-TOFMS and chemometrics to understand potato taste defect. *Microchemical Journal*, *196*, 109578. https://doi.org/10.1016/j.microc.2023.109578

Cain, C. N., Ochoa, G. S., & Synovec, R. E. (2023). Enhancing partial least squares modeling of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data by tile-based variance ranking. *Journal of Chromatography A, 1694*, 463920. https://doi.org/10.1016/j.chroma.2023.463920

Cain, C. N., Trinklein, T. J., Ochoa, G. S., & Synovec, R. E. (2022). Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification. *Analytical Chemistry*, *94*(14), 5658–5666. https://doi.org/10.1021/acs.analchem.2c00223

Cann, A. F., & Liao, J. C. (2008). Production of 2-methyl-1-butanol in engineered Escherichia coli. *Applied Microbiology and Biotechnology*, *81*(1), 89–98. https://doi.org/10.1007/s00253-008-1631-y

Celis Ramírez, A. M., Amézquita, A., Cardona Jaramillo, J. E. C., Matiz-Cerón, L. F., Andrade-Martínez, J. S., Triana, S., Mantilla, M. J., Restrepo, S., Barrios, A. F. G., & de Cock, H. (2020). Analysis of Malassezia Lipidome Disclosed Differences Among the Species and Reveals Presence of Unusual Yeast Lipids. *Frontiers in Cellular and Infection Microbiology*, *10*, 338. https://doi.org/10.3389/fcimb.2020.00338

Chen, C., Zhou, W., Yu, H., Yuan, J., Tian, H. (2020). Evaluation of the perceptual interactions among aldehydes in a cheddar cheese matrix according to odor threshold and aroma intensity. *Molecules, 25*(18), 4308. https://doi.org/10.3390/molecules25184308

Cheung, C., Baker, J. D., Byrne, J. M., & Perrault, K. A. (2022). Investigating volatiles as the secondary metabolome of *Piper methysticum* from root powder and water extracts using comprehensive two-dimensional gas chromatography. *Journal of Ethnopharmacology*, *294*, 115346. https://doi.org/10.1016/j.jep.2022.115346

Chiu, H.-H., & Kuo, C.-H. (2020). Gas chromatography-mass spectrometry-based analytical strategies for fatty acid analysis in biological samples. *Journal of Food and Drug Analysis*, *28*(1), 60–73. https://doi.org/10.1016/j.jfda.2019.10.003

Collins, D. A., Nesterenko, E. P., & Paull, B. (2014). Porous layer open tubular columns in capillary liquid chromatography. *The Analyst*, *139*(6), 1292–1302. https://doi.org/10.1039/C3AN01869E

Contis, E. T., Ho, C.-T., Mussinan, C. J., Parliment, T. H., Shahidi, F., & Spanier, A. M. (1998). *Food Flavors: Formation, Analysis and Packaging Influences*. Elsevier.

Cook-Botelho, J. C., Bachmann, L. M., & French, D. (2017). Chapter 10 - Steroid hormones. In H. Nair & W. Clarke (Eds.), *Mass spectrometry for the clinical laboratory* (pp. 205–230). Academic Press. https://doi.org/10.1016/B978-0-12-800871-3.00010-9

Cordero, C., Schmarr, H.-G., Reichenbach, S. E., & Bicchi, C. (2018). Current Developments in Analyzing Food Volatiles by Multidimensional Gas Chromatographic Techniques. *Journal of Agricultural and Food Chemistry*, *66*(10), 2226–2236. https://doi.org/10.1021/acs.jafc.6b04997

Costa, R., Fanali, C., Pennazza, G., Tedone, L., Dugo, L., Santonico, M., Sciarrone, D., Cacciola, F., Cucchiarini, L., Dachà, M., & Mondello, L. (2015). Screening of volatile compounds composition of white truffle during storage by GCxGC-(FID/MS) and gas sensor array analyses. *LWT - Food Science and Technology*, *60*(2, Part 1), 905–913. https://doi.org/10.1016/j.lwt.2014.09.054

de Hoffmann, E., & Stroobant, V. (2007). *Mass spectrometry: Principles and applications*. John Wiley & Sons.

De Nijs, R. C. M. (1981). Analysis of Light Hydrocarbons C,-C5 with Porous Layer Open Tubular Fused Silica Columns of Aluminum Oxide Part 1: The Column. *J. High Resolut. Chromargr.*, 612–615. https://doi.org/10.1002/jhrc.1240041203

de Oliveira, J. B., Carvalho, M. F. M., de Souza, P. P., & de Lourdes Cardeal, Z. (2024). Development and application of novel extraction method to analyze herbal supplements and adulterant determination by GCxGC/Q-TOFMS. *Food Analytical Methods*. https://doi.org/10.1007/s12161-024-02657-y

de Vanssay, E., Capilla, P., Coscia, D., Do, L., Sternberg, R., & Raulin, F. (1993). Gas chromatography of Titan's atmosphere: IV. Analysis of permanent gases in the presence of hydrocarbons and nitriles with a Molsieve PLOT capillary column. *Journal of Chromatography A, 639*(1), 255–259. https://doi.org/10.1016/0021-9673(93)80261-6

Deetae, P., Bonnarme, P., Spinnler, H. E., & Helinck, S. (2007). Production of volatile aroma compounds by bacterial strains isolated from different surface-ripened French cheeses. *Applied Microbiology and Biotechnology*, *76*(5), 1161–1171. https://doi.org/10.1007/s00253-007-1095-5

Dekeirsschieter, J., Stefanuto, P.-H., Brasseur, C., Haubruge, E., & Focant, J.-F. (2012). Enhanced characterization of the smell of death by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC×GC-TOFMS). *PLOS ONE, 7*(5), e39005. https://doi.org/10.1371/journal.pone.0039005

Dias, L. G., Hacke, A., Bergara, S. F., Villela, O. V., Mariutti, L. R. B., & Bragagnolo, N. (2021). Identification of volatiles and odor-active compounds of aromatic rice by OSME analysis and SPME/GC-MS. *Food Research International, 142*, 110206. https://doi.org/10.1016/j.foodres.2021.110206

Doble, P., Sandercock, M., Du Pasquier, E., Petocz, P., Roux, C., & Dawson, M. (2003). Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic Science International*, *132*(1), 26–39. https://doi.org/10.1016/S0379-0738(03)00002-1

Dunn, W. B., Bailey, N. J. C., & Johnson, H. E. (2005). Measuring the metabolome: Current analytical technologies. *Analyst*, *130*(5), 606–625. https://doi.org/10.1039/B418288J

Edwards, M., Mostafa, A., & Górecki, T. (2011). Modulation in comprehensive two-dimensional gas chromatography: 20 years of innovation. *Analytical and Bioanalytical Chemistry*, *401*(8), 2335–2349. https://doi.org/10.1007/s00216-011-5100-6

Favela, K. A., Hartnett, M. J., Janssen, J. A., Vickers, D. W., Schaub, A. J., Spidle, H. A., & Pickens, K. S. (2021). Nontargeted Analysis of Face Masks: Comparison of Manual Curation to Automated GCxGC Processing Tools. *Journal of the American Society for Mass Spectrometry*, *32*(4), 860–871. https://doi.org/10.1021/jasms.0c00318

Feldhausen, J., Bell, D. C., Yang, Z., Faulhaber, C., Boehm, R., & Heyne, J. (2022). Synthetic aromatic kerosene property prediction improvements with isomer specific characterization via GCxGC and vacuum ultraviolet spectroscopy. *Fuel*, *326*. https://doi.org/10.1016/j.fuel.2022.125002

Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics: Methodology and Distribution* (pp. 66–70). Springer. https://doi.org/10.1007/978-1-4612-4380-9_6

Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome Research*, *13*(2), 244–253. https://doi.org/10.1101/gr.234503

Franchina, F. A., Zanella, D., Dejong, T., & Focant, J.-F. (2021). Impact of the adsorbent material on volatile metabolites during in vitro and in vivo bio-sampling. *Talanta*, *222*, 121569. https://doi.org/10.1016/j.talanta.2020.121569

Freye, C. E., & Synovec, R. E. (2016). High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry. *Talanta*, *161*, 675–680. https://doi.org/10.1016/j.talanta.2016.09.002

Freye, C. E., Bahaghighat, H. D., & Synovec, R. E. (2018). Comprehensive two-dimensional gas chromatography using partial modulation via a pulsed flow valve with a short modulation period. *Talanta*, *177*, 142–149. https://doi.org/10.1016/j.talanta.2017.08.095

Freye, C. E., Mu, L., & Synovec, R. E. (2015). High temperature diaphragm valve-based comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1424*, 127–133. https://doi.org/10.1016/j.chroma.2015.10.098

Frysinger, G. S., & Gaines, R. B. (1999). Comprehensive two-dimensional gas chromatography with mass spectrometric detection (GC x GC/MS) applied to the analysis of petroleum. *HRC Journal of High Resolution Chromatography*, *22*(5), 251–255. https://doi.org/10.1002/(SICI)1521-4168(19990501)22:5<251::AID-JHRC251>3.0.CO;2-V

Gill, M. S., Macleod, A. J., & Moreau, M. (1984). Volatile components of cocoa with particular reference to glucosinolate products. *Phytochemistry, 23*(9), 1937–1942. https://doi.org/10.1016/S0031-9422(00)84945-6

Gonzalez, M., Celis, A. M., Guevara-Suarez, M. I., Molina, J., & Carazzone, C. (2019). Yeast Smell Like What They Eat: Analysis of Volatile Organic Compounds of Malassezia furfur in Growth Media Supplemented with Different Lipids. *Molecules (Basel, Switzerland)*, *24*(3), 419. https://doi.org/10.3390/molecules24030419

Gras, R., Luong, J., Carter, V., Sieben, L., & Cortes, H. (2009). Practical method for the measurement of Alkyl mercaptans in natural gas by multi-dimensional gas chromatography, capillary flow technology, and flame ionization detection. *Journal of Chromatography A*, *1216*(14), 2776–2782. https://doi.org/10.1016/j.chroma.2008.09.029

Griffith, J. F., Winniford, W. L., Sun, K., Edam, R., & Luong, J. C. (2012). A reversed-flow differential flow modulator for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1226*, 116–123. https://doi.org/10.1016/j.chroma.2011.11.036

Grishina, Y. V., Vatlin, A. A., Mavletova, D. A., Odorskaya, M. V., Senkovenko, A. M., Ilyasov, R. A., & Danilenko, V. N. (2023). Metabolites potentially determine the high antioxidant properties of

Limosilactobacillus fermentum U-21. *BioTech, 12*(2), 39. https://doi.org/10.3390/biotech12020039
Gruber, B., Weggler, B. A., Jaramillo, R., Murrell, K. A., Piotrowski, P. K., & Dorman, F. L. (2018). Comprehensive two-dimensional gas chromatography in forensic science: A critical review of recent trends. *TrAC Trends in Analytical Chemistry*, *105*, 292–301. https://doi.org/10.1016/j.trac.2018.05.017
Guan, X., Luong, J., Yu, Z., & Jiang, H. (2020). Quasi-Stop-Flow Modulation Strategy for Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *92*(9), 6251–6256. https://doi.org/10.1021/acs.analchem.0c00814
Guan, X., Zhao, Z., Cai, S., Wang, S., & Lu, H. (2019). Analysis of volatile organic compounds using cryogen-free thermal modulation based comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry. *Journal of Chromatography A*, *1587*, 227–238. https://doi.org/10.1016/j.chroma.2018.12.025
Hanifah, A., Firmanto, H., Putri, S. P., & Fukusaki, E. (2022). Unique metabolite profiles of Indonesian cocoa beans from different origins and their correlation with temperature. *Journal of Bioscience and Bioengineering, 134*(2), 125–132. https://doi.org/10.1016/j.jbiosc.2022.05.001
Hantao, L. W., Teixeira, C. A., & Ferreira, V. H. C. (2022). *CDF2MAT Automated SCRIPT to import NETCDF files to MATLAB | RESAMPLING added to correct RESHAPE for non-integer MS acquisition rates in GCxGC-MS data,*.
Hara, T., Futagami, S., Eeltink, S., De Malsche, W., Baron, G. V., & Desmet, G. (2016). Very High Efficiency Porous Silica Layer Open-Tubular Capillary Columns Produced via in-Column Sol–Gel Processing. *Analytical Chemistry*, *88*(20), 10158–10166. https://doi.org/10.1021/acs.analchem.6b02713
Harries, M. E., & Jeerage, K. M. (2021). Preservation of vapor samples on adsorbent alumina capillaries and implications for field sampling. *Journal of Chromatography A*, *1660*, 462670. https://doi.org/10.1016/j.chroma.2021.462670
Hazeena, S. H., Sindhu, R., Pandey, A., & Binod, P. (2020). Lignocellulosic bio-refinery approach for microbial 2,3-butanediol production. *Bioresource Technology, 302*, 122873. https://doi.org/10.1016/j.biortech.2020.122873
Heinrich, H. T. M., Bremer, P. J., Daughney, C. J., & McQuillan, A. J. (2007). Acid−Base Titrations of Functional Groups on the Surface of the Thermophilic Bacterium Anoxybacillus flavithermus: Comparing a Chemical Equilibrium Model with ATR-IR Spectroscopic Data. *Langmuir*, *23*(5), 2731–2740. https://doi.org/10.1021/la062401j
Heshka, N. E., & Hager, D. B. (2015). Measurement of H2S in Crude Oil and Crude Oil Headspace Using Multidimensional Gas Chromatography, Deans Switching and Sulfur-selective Detection. *Journal of Visualized Experiments : JoVE*, *106*, 53416. https://doi.org/10.3791/53416
Higgins Keppler, E. A., Mead, H. L., Barker, B. M., & Bean, H. D. (2021). Life Cycle Dominates the Volatilome Character of Dimorphic Fungus Coccidioides spp. . *mSphere*, *6*(2). https://doi.org/10.1128/msphere.00040-21
Hinshaw, J. V. (2015). Flame ionization: GC's workhorse detector. *LC-GC North America, 33*(6), 470–477. Retrieved September 3, 2024, from https://www.chromatographyonline.com/view/flame-ionization-gcs-workhorse-detector
Hollywood, K., Brison, D. R., & Goodacre, R. (2006). Metabolomics: Current technologies and future trends. *PROTEOMICS*, *6*(17), 4716–4723. https://doi.org/10.1002/pmic.200600106
Holm, T. (1999). Aspects of the mechanism of the flame ionization detector. *Journal of Chromatography A, 842*(1), 221–227. https://doi.org/10.1016/S0021-9673(98)00706-7

Humston, E. M., Dombek, K. M., Tu, B. P., Young, E. T., & Synovec, R. E. (2011). Toward a global analysis of metabolites in regulatory mutants of yeast. *Analytical and Bioanalytical Chemistry, 401*(8), 2387–2402. https://doi.org/10.1007/s00216-011-4800-2

Humston, E. M., Knowles, J. D., McShea, A., & Synovec, R. E. (2010). Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics. *Journal of Chromatography A, 1217*(12), 1963–1970. https://doi.org/10.1016/j.chroma.2010.01.069

Humston-Fulmer, E. M., Alonso, D. E., & Binkley, J. E. (2020). Chapter Six—Improving cannabis differentiation by expanding coverage of the chemical profile with GCxGC-TOFMS. In I. Ferrer & E. M. Thurman (Eds.), *Comprehensive Analytical Chemistry* (Vol. 90, pp. 169–196). Elsevier. https://doi.org/10.1016/bs.coac.2020.04.007

Hyötyläinen, T., Kallio, M., Hartonen, K., Jussila, M., Palonen, S., & Riekkola, M.-L. (2002). Modulator Design for Comprehensive Two-Dimensional Gas Chromatography: Quantitative Analysis of Polyaromatic Hydrocarbons and Polychlorinated Biphenyls. *Analytical Chemistry*, *74*(17), 4441–4446. https://doi.org/10.1021/ac0201528

Ikonen, K. E., Li, W., Ruonakangas, A., Kilaz, G., & Kenttämaa, H. I. (2023). Different hydrocarbon types in jet fuel as determined by GCxGC/FID—An alternative method to D2425. *Energy & Fuels, 37*(9), 6629–6641. https://doi.org/10.1021/acs.energyfuels.3c00774

Izadmanesh, Y., Garreta-Lara, E., Ghasemi, J. B., Lacorte, S., Matamoros, V., & Tauler, R. (2017). Chemometric analysis of comprehensive two dimensional gas chromatography–mass spectrometry metabolomics data. *Journal of Chromatography A*, *1488*, 113–125. https://doi.org/10.1016/j.chroma.2017.01.052

Jenkins, C. L., & Bean, H. D. (2020). Dependence of the Staphylococcal volatilome composition on microbial nutrition. *Metabolites, 10*(9), 347. https://doi.org/10.3390/metabo10090347

Ji, Z., & Chang, I. L. (1996). A new look at light hydrocarbon separations on commercial alumina PLOT columns: Column selectivity and separation. *Journal of High Resolution Chromatography, 19*(1), 32–36. https://doi.org/10.1002/jhrc.1240190106

Jia, X., Yu, P., An, Q., Ren, J., Fan, G., Wei, Z., Li, X., & Pan, S. (2023). Identification of glucosinolates and volatile odor compounds in microwaved radish (Raphanus sativus L.) seeds and the corresponding oils by UPLC-IMS-QTOF-MS and GC × GC-qMS analysis. *Food Research International, 169*, 112873. https://doi.org/10.1016/j.foodres.2023.112873

Jia, Z., Du, C., Zhu, L., & Wang, Z. (2023). Retention performance of alumina porous layer open-tubular column coated with γ–alumina nanoparticles in the highly volatile hydrocarbons separation. *Journal of Chromatography A*, *1687*, 463657. https://doi.org/10.1016/j.chroma.2022.463657

Jirovetz, L., Smith, D., & Buchbauer, G. (2002). Aroma compound analysis of *Eruca sativa* (Brassicaceae) SPME headspace leaf samples using GC, GC−MS, and olfactometry. *Journal of Agricultural and Food Chemistry, 50*(16), 4643–4646. https://doi.org/10.1021/jf020129n

Jonsson, P., Johansson, A. I., Gullberg, J., Trygg, J., A, J., Grung, B., Marklund, S., Sjöström, M., Antti, H., & Moritz, T. (2005). High-Throughput Data Analysis for Detecting and Identifying Differences between Samples in GC/MS-Based Metabolomic Analyses. *Analytical Chemistry*, *77*(17), 5635–5642. https://doi.org/10.1021/ac050601e

Jordán, M., Goodner, K., & Shaw, P. (2001). Volatile components in banana (*Musa acuminata* colla cv. Cavendish) and yellow passion fruit (*Passiflora edulis* Sims / *flavicarpa* Degner) as determined by GC-MS and GC-olfactometry. *Proceedings of the Florida State Horticultural Society, 114*, 153–157.

Kasukawa, T., Sugimoto, M., Hida, A., Minami, Y., Mori, M., Honma, S., Honma, K., Mishima, K., Soga, T., & Ueda, H. R. (2012). Human blood metabolite timetable indicates internal body time. *Proceedings of the National Academy of Sciences*, *109*(37), 15036–15041. https://doi.org/10.1073/pnas.1207768109

Kaur, J., Sun, N., & Hill, J. E. (2023). Comprehensive profiling of terpenes and terpenoids in different cannabis strains using GC×GC-TOFMS. *Separations, 10*(9), 500. https://doi.org/10.3390/separations10090500

Kellogg, J. J., Graf, T. N., Paine, M. F., McCune, J. S., Kvalheim, O. M., Oberlies, N. H., & Cech, N. B. (2017). Comparison of Metabolomics Approaches for Evaluating the Variability of Complex Botanical Preparations: Green Tea (Camellia sinensis) as a Case Study. *Journal of Natural Products*, *80*(5), 1457–1466. https://doi.org/10.1021/acs.jnatprod.6b01156

Keppler, E. A. H., Dyke, M. C. C. V., Mead, H. L., Lake, D. F., Magee, D. M., Barker, B. M., & Bean, H. D. (2022). Volatile metabolites in lavage fluid are correlated to Valley fever disease severity in murine model lung infections. *bioRxiv*. https://doi.org/10.1101/2022.10.02.510537

Kimura, Y., Shibasaki, S., Morisato, K., Ishizuka, N., Minakuchi, H., Nakanishi, K., Matsuo, M., Amachi, T., Ueda, M., & Ueda, K. (2004). Microanalysis for MDR1 ATPase by high-performance liquid chromatography with a titanium dioxide column. *Analytical Biochemistry*, *326*(2), 262–266. https://doi.org/10.1016/j.ab.2003.12.012

Kirkland, J. J. (1963). Fibrillar Boehmite—A new adsorbent for gas solid chromatography. *Analytical Chemistry, 35*(10), 1295–1297. https://doi.org/10.1021/ac60202a038

Klee, M. S., Cochran, J., Merrick, M., & Blumberg, L. M. (2015). Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain. *Journal of Chromatography A*, *1383*, 151–159. https://doi.org/10.1016/j.chroma.2015.01.031

Klevecz, R. R., Bolen, J., Forrest, G., & Murray, D. B. (2004). A genomewide oscillation in transcription gates DNA replication and cell cycle. *Proceedings of the National Academy of Sciences*, *101*(5), 1200–1205. https://doi.org/10.1073/pnas.0306490101

Koljančić, N., Gomes, A. A., & Špánik, I. (2023). A non-target geographical origin screening of botrytized wines through comprehensive two-dimensional gas chromatography coupled with high-resolution mass spectrometry. *Journal of Separation Science, 46*(7), 2300249. https://doi.org/10.1002/jssc.202300249

Koljančić, N., Vyviurska, O., & Špánik, I. (2023). Aroma compounds in essential oils: Analyzing chemical composition using two-dimensional gas chromatography–high resolution time-of-flight mass spectrometry combined with chemometrics. *Plants, 12*(12), 2362. https://doi.org/10.3390/plants12122362

Korytár, P., Parera, J., Leonards, P. E. G., De Boer, J., & Brinkman, U. A. T. (2005). Quadrupole mass spectrometer operating in the electron-capture negative ion mode as detector for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1067*(1–2), 255–264. https://doi.org/10.1016/j.chroma.2004.11.045

Kotnik, P., Čolnik, M., Finšgar, M., Knez, Ž., & Škerget, M. (2020). Determination of C1–C6 hydrocarbons in gaseous plastic degradation products by GC–MS method. *Polymer Degradation and Stability, 182*, 109386. https://doi.org/10.1016/j.polymdegradstab.2020.109386

Krupčík, J., Májek, P., Gorovenko, R., Špánik, I., Sandra, P., & Armstrong, D. W. (2013). On the determination of a detector response enhancement factor for flow modulated comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1286*, 235–240. https://doi.org/10.1016/j.chroma.2013.02.068

Kulsing, C., Knob, R., Macka, M., Junor, P., Boysen, R. I., & Hearn, M. T. W. (2014). Molecular imprinted polymeric porous layers in open tubular capillaries for chiral separations. *Journal of Chromatography A*, *1354*, 85–91. https://doi.org/10.1016/j.chroma.2014.05.065

Lebedev, A. T., Polyakova, O. V., Artaev, V. B., Mednikova, M. B., & Anokhina, E. A. (2021). Comprehensive two-dimensional gas chromatography-high resolution mass spectrometry with complementary ionization methods in the study of 5000-year-old mummy. *Rapid Communications in Mass Spectrometry*, *35*(8). https://doi.org/10.1002/rcm.9058

LECO Corporation. (n.d.). *ChromaTOF Tile analytical software*. Retrieved September 4, 2024, from https://www.leco.com/product/chromatof-tile

LECO Corporation. (n.d.). Thermal modulator. Retrieved September 3, 2024, from https://www.leco.com/products/thermal-modulator/

Lee, A. L., Bartle, K. D., & Lewis, A. C. (2001). A Model of Peak Amplitude Enhancement in Orthogonal Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *73*(6), 1330–1335. https://doi.org/10.1021/ac001120s

Liao, W. C., Ou-Yang, C. F., Wang, C. H., Chang, C. C., & Wang, J. L. (2013). Two-dimensional gas chromatographic analysis of ambient light hydrocarbons. *Journal of Chromatography A*, *1294*, 122–129. https://doi.org/10.1016/j.chroma.2013.04.008

Lill, J. (2003). Proteomic tools for quantitation by mass spectrometry. *Mass Spectrometry Reviews, 22*(3), 182–194. https://doi.org/10.1002/mas.10048

Lin, L.-J., Saini, M., Chiang, C.-J., & Chao, Y.-P. (2021). Biocatalytic Conversion of Short-Chain Fatty Acids to Corresponding Alcohols in Escherichia coli. *Processes*, *9*(6), Article 6. https://doi.org/10.3390/pr9060973

Liu, X., Sun, S., Nie, R., Ma, J., Qu, Q., & Yang, L. (2018). Highly uniform porous silica layer open-tubular capillary columns produced via in-situ biphasic sol–Gel processing for open-tubular capillary electrochromatography. *Journal of Chromatography A*, *1538*, 86–93. https://doi.org/10.1016/j.chroma.2018.01.024

Liu, Z., & Phillips, J. B. (1991). Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface. *J. Chromatogr. Sci.*, *29*, 227–231. https://doi.org/10.1093/chromsci/29.6.227

Lovestead, T. M., & Bruno, T. J. (2010). Trace Headspace Sampling for Quantitative Analysis of Explosives with Cryoadsorption on Short Alumina Porous Layer Open Tubular Columns. *Analytical Chemistry*, *82*(13), 5621–5627. https://doi.org/10.1021/ac1005926

Lübeck, J. S., Alexandrino, G. L., & Christensen, J. H. (2020). GC × GC–HRMS nontarget fingerprinting of organic micropollutants in urban freshwater sediments. *Environmental Sciences Europe, 32*(1), 78. https://doi.org/10.1186/s12302-020-00353-2

Luo, Q., Rejtar, T., Wu, S.-L., & Karger, B. L. (2009). Hydrophilic interaction 10μm I.D. porous layer open tubular columns for ultratrace glycan analysis by liquid chromatography–mass spectrometry. *Journal of Chromatography A*, *1216*(8), 1223–1231. https://doi.org/10.1016/j.chroma.2008.09.105

Luong, J., Guan, X., Xu, S., Gras, R., & Shellie, R. A. (2016). Thermal Independent Modulator for Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *88*(17), 8428–8432. https://doi.org/10.1021/acs.analchem.6b02525

Makhubela, P. C. K., Rohwer, E. R., & Naudé, Y. (2023). Detection of tuberculosis-associated compounds from human skin by GCxGC-TOFMS. *Journal of Chromatography B*, *1231*, 123937. https://doi.org/10.1016/j.jchromb.2023.123937

Mani-Varnosfaderani, A., Gao, A., Poch, K. R., Caceres, S. M., Nick, J. A., & Hill, J. E. (2022). Breath biomarkers associated with nontuberculosis mycobacteria disease status in persons with cystic fibrosis: A pilot study. *Journal of Breath Research*, *16*(3). https://doi.org/10.1088/1752-7163/ac6bb6

Mara, P., Nelson, R. K., Reddy, C. M., Teske, A., & Edgcomb, V. P. (2022). Sterane and hopane biomarkers capture microbial transformations of complex hydrocarbons in young hydrothermal Guaymas Basin sediments. *Communications Earth & Environment, 3*(1), 1–14. https://doi.org/10.1038/s43247-022-00582-8

Marney, L. C., Christopher Siegler, W., Parsons, B. A., Hoggard, J. C., Wright, B. W., & Synovec, R. E. (2013). Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry data. *Talanta*, *115*, 887–895. https://doi.org/10.1016/j.talanta.2013.06.038

Marriott, P. J., Chin, S.-T., & Nolvachai, Y. (2021). Techniques and application in comprehensive multidimensional gas chromatography – mass spectrometry. *Journal of Chromatography A*, *1636*, 461788. https://doi.org/10.1016/j.chroma.2020.461788

Marseglia, A., Musci, M., Rinaldi, M., Palla, G., & Caligiani, A. (2020). Volatile fingerprint of unroasted and roasted cocoa beans (*Theobroma cacao* L.) from different geographical origins. *Food Research International, 132*, 109101. https://doi.org/10.1016/j.foodres.2020.109101

Martín, A., Benito, M. J., Aranda, E., Ruiz-Moyano, S., Córdoba, J. J., & Córdoba, M. G. (2010). Characterization by Volatile Compounds of Microbial Deep Spoilage in Iberian Dry-Cured Ham. *Journal of Food Science*, *75*(6), M360–M365. https://doi.org/10.1111/j.1750-3841.2010.01674.x

Maštovská, K., Hajšlová, J., & Lehotay, S. J. (2004). Ruggedness and other performance characteristics of low-pressure gas chromatography–mass spectrometry for the fast analysis of multiple pesticide residues in food crops. *Journal of Chromatography A*, *1054*(1), 335–349. https://doi.org/10.1016/j.chroma.2004.08.061

Mazur, D. M., Sosnova, A. A., Latkin, T. B., Artaev, B. V., Siek, K., Koluntaev, D. A., & Lebedev, A. T. (2023). Application of clusterization algorithms for analysis of semivolatile pollutants in Arkhangelsk snow. *Analytical and Bioanalytical Chemistry, 415*(9), 2587–2599. https://doi.org/10.1007/s00216-022-04390-z

Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., McCrindle, R., & Atlabachew, M. (2019). GC-MS profiling of fatty acids in green coffee (*Coffea arabica* L.) beans and chemometric modeling for tracing geographical origins from Ethiopia. *Journal of the Science of Food and Agriculture, 99*(9), 3811–3823. https://doi.org/10.1002/jsfa.9603

Melder, J., Zinsmeister, J., Grein, T., Jürgens, S., Köhler, M., & Oßwald, P. (2023). Comprehensive two-dimensional gas chromatography: A universal method for composition-based prediction of emission characteristics of complex fuels. *Energy Fuels, 37*(9), 4580–4595. https://doi.org/10.1021/acs.energyfuels.2c04270

Menéndez-Carreño, M., Steenbergen, H., & Janssen, H.-G. (2012). Development and validation of a comprehensive two-dimensional gas chromatography-mass spectrometry method for the analysis of phytosterol oxidation products in human plasma. *Analytical and Bioanalytical Chemistry*, *402*(6), 2023–2032. https://doi.org/10.1007/s00216-011-5432-2

Mikaliunaite, L., & Synovec, R. E. (2022). Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Talanta*, *244*, 123396. https://doi.org/10.1016/j.talanta.2022.123396

Mikaliunaite, L., & Synovec, R. E. (2024). Simultaneous discovery of compounds dominated by either molding kinetics or geographical region of origin for moisture damaged cacao beans using

orthogonally applied tile-based fisher ratio analysis of GC×GC-TOFMS data. *Journal of Chromatography A*, *1730*, 465093. https://doi.org/10.1016/j.chroma.2024.465093

Mikaliunaite, L., Sudol, P. E., Cain, C. N., & Synovec, R. E. (2021). Baseline correction method for dynamic pressure gradient modulated comprehensive two-dimensional gas chromatography with flame ionization detection. *Journal of Chromatography A*, *1652*, 462358. https://doi.org/10.1016/j.chroma.2021.462358

Mikaliunaite, L., Trinklein, T. J., Ochoa, G. S., Sudol, P. E., Bell, D. S., & Synovec, R. E. (2023). Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension. *Journal of Chromatography Open*, *3*, 100076. https://doi.org/10.1016/j.jcoa.2022.100076

Mishra, G., Gupta, P., Kumar, A., Purohit, A., Tak, V., & Pardasani, D. (2024). Gas chromatography tandem mass spectrometric analysis of alkylphosphonofluoridic acids as verification targets of nerve agents. *Journal of Chromatography A*, *1716*, 464645. https://doi.org/10.1016/j.chroma.2024.464645

Moco, S., Schneider, B., & Vervoort, J. (2009). Plant Micrometabolomics: The Analysis of Endogenous Metabolites Present in a Plant Cell or Tissue. *Journal of Proteome Research*, *8*(4), 1694–1703. https://doi.org/10.1021/pr800973r

Mohamadi Alasti, F., Asefi, N., Maleki, R., & SeiiedlouHeris, S. S. (2019). Investigating the flavor compounds in the cocoa powder production process. *Food Science & Nutrition, 7*(11), 3892–3901. https://doi.org/10.1002/fsn3.1244

Mohler, R. E., Dombek, K. M., Hoggard, J. C., Pierce, K. M., Young, E. T., & Synovec, R. E. (2007). Comprehensive analysis of yeast metabolite GC×GC–TOFMS data: Combining discovery-mode and deconvolution chemometric software. *Analyst, 132*(8), 756–767. https://doi.org/10.1039/B700061H

Mohler, R. E., Dombek, K. M., Hoggard, J. C., Young, E. T., & Synovec, R. E. (2006). Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Analysis of Metabolites in Fermenting and Respiring Yeast Cells. *Analytical Chemistry*, *78*(8), 2700–2709. https://doi.org/10.1021/ac052106o

Mohler, R. E., Prazen, B. J., & Synovec, R. E. (2006). Total-transfer, valve-based comprehensive two-dimensional gas chromatography. *Analytica Chimica Acta*, *555*(1), 68–74. https://doi.org/10.1016/j.aca.2005.08.072

Mohler, R. E., Tu, B. P., Dombek, K. M., Hoggard, J. C., Young, E. T., & Synovec, R. E. (2008). Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography–time-of-flight-mass spectrometry data. *Journal of Chromatography A*, *1186*(1), 401–411. https://doi.org/10.1016/j.chroma.2007.10.063

Mondello, L., Casilli, A., Tranchida, P. Q., Dugo, G., & Dugo, P. (2005). Comprehensive two-dimensional gas chromatography in combination with rapid scanning quadrupole mass spectrometry in perfume analysis. *Journal of Chromatography A*, *1067*(1), 235–243. https://doi.org/10.1016/j.chroma.2004.09.040

Moreira, L. S., d'Avila, L. A., & Azevedo, D. A. (2003). Automotive Gasoline Quality Analysis by Gas Chromatography: Study of Adulteration. *Chromatographia*, *58*(7), 501–505. https://doi.org/10.1365/s10337-003-0065-z

Mostafa, A., & Górecki, T. (2016). Development and Design of a Single-Stage Cryogenic Modulator for Comprehensive Two-Dimensional Gas Chromatography. *Analytical Chemistry*, *88*(10), 5414–5423. https://doi.org/10.1021/acs.analchem.6b00767

Mucédola, V., Vieira, L. C. S., Pierone, D., Gobbi, A. L., Poppi, R. J., & Hantao, L. W. (2017). Thermal desorption modulation for comprehensive two-dimensional gas chromatography using a simple and inexpensive segmented-loop fluidic interface. *Talanta*, *164*, 470–476. https://doi.org/10.1016/j.talanta.2016.12.005

Nelson, R. K., Forsythe, J., Eiserbeck, C., Scarlett, A. G., Grice, K., Mullins, O. C., & Reddy, C. M. (2022). GC×GC analysis of novel 2α-methyl biomarker compounds from a large Middle East oilfield. *Energy Fuels, 36*(10), 8853–8865. https://doi.org/10.1021/acs.energyfuels.2c00949

Nichols, J. E., Harries, M. E., Lovestead, T. M., & Bruno, T. J. (2014). Analysis of arson fire debris by low temperature dynamic headspace adsorption porous layer open tubular columns. *Journal of Chromatography A*, *1334*, 126–138. https://doi.org/10.1016/j.chroma.2014.01.080

Nolvachai, Y., Kulsing, C., & Marriott, P. J. (2017). Multidimensional gas chromatography in food analysis. *TrAC Trends in Analytical Chemistry*, *96*, 124–137. https://doi.org/10.1016/j.trac.2017.05.001

Nouri, N., Sun, N., & Hill, J. E. (2024). A feasibility study of sample re-collection in the analysis of selected volatile compounds in breath samples using GC×GC-TOFMS. *Journal of Chromatography A, 1730*, 465125. https://doi.org/10.1016/j.chroma.2024.465125

Nyangale, Edna. P., Mottram, Donald. S., & Gibson, Glenn. R. (2012). Gut Microbial Activity, Implications for Health and Disease: The Potential Role of Metabolite Analysis. *Journal of Proteome Research*, *11*(12), 5573–5585. https://doi.org/10.1021/pr300637d

Ochoa, G. S., & Synovec, R. E. (2023). Investigating analyte breakthrough under non-linear isotherm conditions during solid phase extraction facilitated by non-targeted analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Talanta, 259*, 124525. https://doi.org/10.1016/j.talanta.2023.124525

Ochoa, G. S., Billingsley, M. C., & Synovec, R. E. (2023). Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: Comparative analysis of aerospace fuel composition. *Analytical and Bioanalytical Chemistry*, *415*(13), 2411–2423. https://doi.org/10.1007/s00216-022-04348-1

Ochoa, G. S., Prebihalo, S. E., Reaser, B. C., Marney, L. C., & Synovec, R. E. (2020). Statistical inference of mass channel purity from Fisher ratio analysis using comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry data. *Journal of Chromatography A, 1627*, 461401. https://doi.org/10.1016/j.chroma.2020.461401

Oldridge, N., Panic, O., & Górecki, T. (2008). Stop-flow comprehensive two-dimensional gas chromatography with pneumatic switching. *Journal of Separation Science*, *31*(19), 3375–3384. https://doi.org/10.1002/jssc.200800265

Oliva, D. J., Cheung, C., & Perrault, K. A. (2023). Fisher ratio feature selection by manual peak area calculations on comprehensive two-dimensional gas chromatography data using standard mixtures with variable composition, storage, and interferences. *Analytical and Bioanalytical Chemistry, 415*(8), 2575–2585. https://doi.org/10.1007/s00216-022-04484-8

Otaif, K. D., Badjah-Hadj-Ahmed, A.-Y., & ALOthman, Z. A. (2024). Preparation of UiO-66 MOF-Bonded Porous-Layer Open-Tubular Columns Using an In Situ Growth Approach for Gas Chromatography. *Molecules*, *29*(11), Article 11. https://doi.org/10.3390/molecules29112505

Ou-Yang, C.-F., Chen, Y.-J., Hsieh, H.-C., Lee, C.-T., Chi, K.-H., Lin, N.-H., Chang, C.-C., & Wang, J.-L. (2024). Identification of organic constituents on atmospheric particulate matter in the East Asian background air of free troposphere by GC×GC-TOFMS. *Chemosphere, 364*, 143095. https://doi.org/10.1016/j.chemosphere.2024.143095

Parsons, B. A., Marney, L. C., Siegler, W. C., Hoggard, J. C., Wright, B. W., & Synovec, R. E. (2015). Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach. *Analytical Chemistry*, *87*(7), 3812–3819. https://doi.org/10.1021/ac504472s

Parsons, B. A., Pinkerton, D. K., & Synovec, R. E. (2018). Implications of phase ratio for maximizing peak capacity in comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Journal of Chromatography A*, *1536*, 16–26. https://doi.org/10.1016/j.chroma.2017.07.018

Patrushev, Y. V., & Sidelnikov, V. N. (2018). Selection of the porous layer open tubular columns for separation of light components in comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1579*, 83–88. https://doi.org/10.1016/j.chroma.2018.10.015

Peano, A., Johnson, E., Chiavassa, E., Tizzani, P., Guillot, J., & Pasquetti, M. (2020). Antifungal Resistance Regarding Malassezia pachydermatis: Where Are We Now? *Journal of Fungi*, *6*(2), 93. https://doi.org/10.3390/jof6020093

Pereira, G. F. L., Fileti, E. E., & Siqueira, L. J. A. (2021). Comparing Graphite and Graphene Oxide Supercapacitors with a Constant Potential Model. *The Journal of Physical Chemistry C*, *125*(4), 2318–2326. https://doi.org/10.1021/acs.jpcc.0c10347

Peroni, D., Vonk, R. J., van Egmond, W., & Janssen, H.-G. (2012). Macroporous polymer monoliths as second dimension columns in comprehensive two-dimensional gas chromatography: A feasibility study. *Journal of Chromatography A, 1268*, 139–149. https://doi.org/10.1016/j.chroma.2012.10.019

Pesesse, R., Stefanuto, P.-H., Schleich, F., Louis, R., & Focant, J.-F. (2019). Multimodal chemometric approach for the analysis of human exhaled breath in lung cancer patients by TD-GC×GC-TOFMS. *Journal of Chromatography B, 1114–1115*, 146–153. https://doi.org/10.1016/j.jchromb.2019.01.029

Petculescu, A., & Lueptow, R. M. (2012). Quantitative Acoustic Relaxational Spectroscopy for real-time monitoring of natural gas: A perspective on its potential. *Sensors and Actuators B: Chemical*, *169*, 121–127. https://doi.org/10.1016/j.snb.2012.03.086

Peterson, R. J., & Chang, S. S. (1982). Identification of volatile flavor compounds of fresh, frozen beef stew and a comparison of these with those of canned beef stew. *Journal of Food Science, 47*(4), 1444–1448. https://doi.org/10.1111/j.1365-2621.1982.tb04957.x

Petitjean, D. L., & Leftault, C. J. (1963). Oxide-coated aluminum tubing for capillary gas chromatography. *Journal of Chromatographic Science, 1*(3), 18–21. https://doi.org/10.1093/chromsci/1.3.18

Pichersky, E., & Gang, D. R. (2000). Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective. *Trends in Plant Science*, *5*(10), 439–445. https://doi.org/10.1016/S1360-1385(00)01741-6

Pierce, K. M., Hoggard, J. C., Hope, J. L., Rainey, P. M., Hoofnagle, A. N., Jack, R. M., Wright, B. W., & Synovec, R. E. (2006). Fisher ratio method applied to third-order separation data to identify significant chemical components of metabolite extracts. *Analytical Chemistry, 78*(14), 5068–5075. https://doi.org/10.1021/ac0602625

Pierce, K. M., Wood, L. F., Wright, B. W., & Synovec, R. E. (2005). A Comprehensive Two-Dimensional Retention Time Alignment Algorithm To Enhance Chemometric Analysis of Comprehensive Two-Dimensional Separation Data. *Analytical Chemistry*, *77*(23), 7735–7743. https://doi.org/10.1021/ac0511142

Pratt, G. C., Wu, C. Y., Bock, D., Adgate, J. L., Ramachandran, G., Stock, T. H., Morandi, M., & Sexton, K. (2004). Comparing Air Dispersion Model Predictions with Measured Concentrations of VOCs in Urban Communities. *Environmental Science & Technology*, *38*(7), 1949–1959. https://doi.org/10.1021/es0306381

Prebihalo, S. E., Berrier, K. L., Freye, C. E., Bahaghighat, H. D., Moore, N. R., Pinkerton, D. K., & Synovec, R. E. (2018). Multidimensional Gas Chromatography: Advances in Instrumentation, Chemometrics, and Applications. *Analytical Chemistry*, *90*(1), 505–532. https://doi.org/10.1021/acs.analchem.7b04226

Prebihalo, S. E., Ochoa, G. S., Berrier, K. L., Skogerboe, K. J., Cameron, K. L., Trump, J. R., Svoboda, S. J., Wickiser, J. K., & Synovec, R. E. (2020). Control-Normalized Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data for Enhanced Biomarker Discovery in a Metabolomic Study of Orthopedic Knee-Ligament Injury. *Analytical Chemistry*, *92*(23), 15526–15533. https://doi.org/10.1021/acs.analchem.0c03456

PubChem. (n.d.). *2,3-Dimethyl-2-cyclopenten-1-one*. Retrieved July 18, 2024, from https://pubchem.ncbi.nlm.nih.gov/compound/14270

Purcaro, G., Tranchida, P. Q., Ragonese, C., Conte, L., Dugo, P., Dugo, G., & Mondello, L. (2010). Evaluation of a rapid-scanning quadrupole mass spectrometer in an apolar × ionic-liquid comprehensive two-dimensional gas chromatography system. *Analytical Chemistry*, *82*(20), 8583–8590. https://doi.org/10.1021/ac101678r

Rahman, M. M., El-Aty, A. M. A., Choi, J.-H., Shin, H.-C., Shin, S. C., & Shim, J.-H. (2015). Basic Overview on Gas Chromatography Columns. In *Analytical Separation Science* (pp. 823–834). American Cancer Society. https://doi.org/10.1002/9783527678129.assep024

Raninen, K., Nenonen, R., Järvelä-Reijonen, E., Poutanen, K., Mykkänen, H., & Raatikainen, O. (2021). Comprehensive Two-Dimensional Gas Chromatography–Mass Spectrometry Analysis of Exhaled Breath Compounds after Whole Grain Diets. *Molecules*, *26*(9), Article 9. https://doi.org/10.3390/molecules26092667

Rasheed, D. M., Serag, A., Abdel Shakour, Z. T., & Farag, M. (2021). Novel trends and applications of multidimensional chromatography in the analysis of food, cosmetics and medicine bearing essential oils. In *Talanta* (Vol. 223). Elsevier B.V. https://doi.org/10.1016/j.talanta.2020.121710

Ratiu, I.-A., Ligor, T., Bocos-Bintintan, V., Al-Suod, H., Kowalkowski, T., Rafińska, K., & Buszewski, B. (2017). The effect of growth medium on an Escherichia coli pathway mirrored into GC/MS profiles. *Journal of Breath Research*, *11*(3), 036012. https://doi.org/10.1088/1752-7163/aa7ba2

Reid III, G. L., & Armstrong, D. W. (1994). Cyclodextrin PLOT columns for the gas-solid chromatographic separation of light hydrocarbons and inorganic gases. *Journal of Microcolumn Separations, 6*(3), 151–157. https://doi.org/10.1002/mcs.1220060208

Restek. (2020). Restek's PLOT column family. Retrieved July 20, 2022, from https://www.restek.com/en/technical-literature-library/articles/resteks-PLOT-column-family/

Ribeiro, C., Gonçalves, R., & Tiritan, M. E. (2021). Separation of Enantiomers Using Gas Chromatography: Application in Forensic Toxicology, Food and Environmental Analysis. *Critical Reviews in Analytical Chemistry*, *51*(8), 787–811. https://doi.org/10.1080/10408347.2020.1777522

Rios-Navarro, A., Gonzalez, M., Carazzone, C., & Celis Ramírez, A. M. (2021). Learning about microbial language: Possible interactions mediated by microbial volatile organic compounds (VOCs) and relevance to understanding Malassezia spp. metabolism. *Metabolomics*, *17*(4), 39. https://doi.org/10.1007/s11306-021-01786-3

Rios-Navarro, A., Gonzalez, M., Carazzone, C., & Celis Ramírez, A. M. (2023). Why Do These Yeasts Smell So Good? Volatile Organic Compounds (VOCs) Produced by Malassezia Species in the Exponential and Stationary Growth Phases. *Molecules*, *28*(6), 2620. https://doi.org/10.3390/molecules28062620

Rodriguez-Campos, J., Escalona-Buendía, H. B., Orozco-Avila, I., Lugo-Cervantes, E., & Jaramillo-Flores, M. E. (2011). Dynamics of volatile and non-volatile compounds in cocoa (*Theobroma cacao* L.) during fermentation and drying processes using principal components analysis. *Food Research International, 44*(1), 250–258. https://doi.org/10.1016/j.foodres.2010.10.028

Rosso, M. C., Mazzucotelli, M., Bicchi, C., Charron, M., Manini, F., Menta, R., Fontana, M., Reichenbach, S. E., & Cordero, C. (2020). Adding extra dimensions to hazelnuts primary metabolome fingerprinting by comprehensive two-dimensional gas chromatography combined with time-of-flight mass spectrometry featuring tandem ionization: Insights on the aroma potential. *Journal of Chromatography A, 1614*, 460739. https://doi.org/10.1016/j.chroma.2019.460739

Safarova, V. I., Sapelnikova, S. V., Djazhenko, E. V., Teplova, G. I., Shajdulina, G. F., & Kudasheva, F. K. (2004). Gas chromatography–mass spectrometry with headspace for the analysis of volatile organic compounds in waste water. *Journal of Chromatography B*, *800*(1), 325–330. https://doi.org/10.1016/j.jchromb.2003.10.070

Sastoque, A., Triana, S., Ehemann, K., Suarez, L., Restrepo, S., Wösten, H., de Cock, H., Fernández-Niño, M., González Barrios, A. F., & Celis Ramírez, A. M. (2020). New Therapeutic Candidates for the Treatment of Malassezia pachydermatis -Associated Infections. *Scientific Reports*, *10*(1), 4860. https://doi.org/10.1038/s41598-020-61729-1

Sawoszczuk, T., Syguła-Cholewińska, J., & del Hoyo-Meléndez, J. M. (2015). Optimization of headspace solid phase microextraction for the analysis of microbial volatile organic compounds emitted by fungi: Application to historical objects. *Journal of Chromatography A, 1409*, 30–45. https://doi.org/10.1016/j.chroma.2015.07.059

Schöneich, S., Gough, D. V., Trinklein, T. J., & Synovec, R. E. (2020). Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry detection. *Journal of Chromatography A*, *1620*, 460982. https://doi.org/10.1016/j.chroma.2020.460982

Schöneich, S., Ochoa, G. S., Monzón, C. M., & Synovec, R. E. (2022). Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography / mass spectrometry data: Study of the pacu fish metabolome. *Journal of Chromatography A, 1667*, 462868. https://doi.org/10.1016/j.chroma.2022.462868

Schöneich, S., Trinklein, T. J., Warren, C. G., & Synovec, R. E. (2020). A systematic investigation of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry with dynamic pressure gradient modulation for high peak capacity separations. *Analytica Chimica Acta*, *1134*, 115–124. https://doi.org/10.1016/j.aca.2020.08.023

Schwan, R. F., Bressani, A. P. P., Martinez, S. J., Batista, N. N., & Dias, D. R. (2023). The essential role of spontaneous and starter yeasts in cocoa and coffee fermentation. *FEMS Yeast Research*, *23*, foad019. https://doi.org/10.1093/femsyr/foad019

Seeley, J. V., Kramp, F., & Hicks, C. J. (2000). Comprehensive Two-Dimensional Gas Chromatography via Differential Flow Modulation. *Analytical Chemistry*, *72*(18), 4346–4352. https://doi.org/10.1021/ac000249z

Sekiguchi, Y., Mitsuhashi, N., Inoue, Y., Yagisawa, H., & Mimura, T. (2004). Analysis of sugar phosphates in plants by ion chromatography on a titanium dioxide column with pulsed

amperometric detection. *Journal of Chromatography A*, *1039*(1), 71–76. https://doi.org/10.1016/j.chroma.2004.02.015
Serrano, A., & Gallego, M. (2006). Sorption study of 25 volatile organic compounds in several Mediterranean soils using headspace–gas chromatography–mass spectrometry. *Journal of Chromatography A*, *1118*(2), 261–270. https://doi.org/10.1016/j.chroma.2006.03.095
Sheibani, E., Duncan, S. E., Kuhn, D. D., Dietrich, A. M., Newkirk, J. J., & O'Keefe, S. F. (2016). Changes in flavor volatile composition of oolong tea after panning during tea processing. *Food Science & Nutrition, 4*(3), 456–468. https://doi.org/10.1002/fsn3.307
Shellie, R. A., & Marriott, P. J. (2003). Comprehensive two-dimensional gas chromatography-mass spectrometry analysis of Pelargonium graveolens essential oil using rapid scanning quadrupole mass spectrometry. *Analyst*, *128*(7), 879–883. https://doi.org/10.1039/b304371a
Shellie, R. A., Cortes, H. J., Gras, R., & Luong, J. (2013). Planar microfluidic devices in flow modulated comprehensive two dimensional gas chromatography for challenging petrochemical applications. *Analytical Methods*, *5*(23), 6598–6604. https://doi.org/10.1039/c3ay41198b
Shi, X., Zhao, S., Chen, S., Han, X., Yang, Q., Zhang, L., Xia, X., Tu, J., & Hu, Y. (2022). Tetramethylpyrazine in Chinese baijiu: Presence, analysis, formation, and regulation. *Frontiers in Nutrition, 9*. https://doi.org/10.3389/fnut.2022.1004435 (accessed December 22, 2023)
Sholokhova, A. Y., Patrushev, Y. V., Sidelnikov, V. N., & Buryak, A. K. (2020). Analysis of light components in pyrolysis products by comprehensive two-dimensional gas chromatography with PLOT columns. *Talanta*, *209*, 120448. https://doi.org/10.1016/j.talanta.2019.120448
Skoczyńska, E., Korytár, P., & Boer, J. de. (2008). Maximizing Chromatographic Information from Environmental Extracts by GCxGC-ToF-MS. *Environmental Science & Technology*, *42*(17), 6611–6618. https://doi.org/10.1021/es703229t
Song, C., Zhang, Y., Zhao, Q., Chen, M., Zhang, Y., Gao, C., Jia, Z., Song, S., Guan, J., & Shang, Z. (2024). Volatile organic compounds produced by *Bacillus aryabhattai* AYG1023 against *Penicillium expansum* causing blue mold on the Huangguan pear. *Microbiological Research, 278*, 127531. https://doi.org/10.1016/j.micres.2023.127531
Stefanuto, P. H., Romano, R., Rees, C. A., Nasir, M., Thakuria, L., Simon, A., Reed, A. K., Marczin, N., & Hill, J. E. (2022). Volatile organic compound profiling to explore primary graft dysfunction after lung transplantation. *Scientific Reports*, *12*, 2053. https://doi.org/10.1038/s41598-022-05994-2
Stefanuto, P.-H., Smolinska, A., & Focant, J.-F. (2021). Advanced chemometric and data handling tools for GC×GC-TOF-MS: Application of chemometrics and related advanced data handling in chemical separations. *Trends in Analytical Chemistry, 139*, 116251. https://doi.org/10.1016/j.trac.2021.116251
Stilo, F., Gabetti, E., Bicchi, C., Carretta, A., Peroni, D., Reichenbach, S. E., Cordero, C., & McCurry, J. (2020). A step forward in the equivalence between thermal and differential-flow modulated comprehensive two-dimensional gas chromatography methods. *Journal of Chromatography A*, *1627*, 461396. https://doi.org/10.1016/j.chroma.2020.461396
Stufkens, J. S., & Bogaard, H. J. (1975). Rapid method for the determination of the composition of natural gas by gas chromatography. *Analytical Chemistry*, *47*(3), 383–386. https://doi.org/10.1021/ac60353a060
Su, H., Chen, H., & Lin, J. (2020). Enriching the Production of 2-Methyl-1-Butanol in Fermentation Process Using Corynebacterium crenatum. *Current Microbiology*, *77*(8), 1699–1706. https://doi.org/10.1007/s00284-020-01961-0

Sudol, P. E., Galletta, M., Tranchida, P. Q., Zoccali, M., Mondello, L., & Synovec, R. E. (2022). Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis. *Journal of Chromatography A*, *1662*, 462735. https://doi.org/10.1016/j.chroma.2021.462735

Sudol, P. E., Ochoa, G. S., & Synovec, R. E. (2021). Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Journal of Chromatography A*, *1644*, 462092. https://doi.org/10.1016/j.chroma.2021.462092

Sudol, P. E., Ochoa, G. S., Cain, C. N., & Synovec, R. E. (2022). Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry. *Analytica Chimica Acta*, *1209*, 339847. https://doi.org/10.1016/j.aca.2022.339847

Sudol, P. E., Pierce, K. M., Prebihalo, S. E., Skogerboe, K. J., Wright, B. W., & Synovec, R. E. (2020). Development of gas chromatographic pattern recognition and classification tools for compliance and forensic analyses of fuels: A review. *Analytica Chimica Acta*, *1132*, 157–186. https://doi.org/10.1016/j.aca.2020.07.027

Sun, N., Krishnan, P., Rees, C. A., Zhang, M., Stevenson, K. A. J. M., & Hill, J. E. (2023). Profiling volatile organic compounds from human plasma using GC × GC-ToFMS. *Journal of Breath Research, 17*(3), 037104. https://doi.org/10.1088/1752-7163/acd806

Syrokou, M. K., Paramithiotis, S., Kanakis, C. D., Papadopoulos, G. K., Tarantilis, P. A., Skandamis, P. N., Bosnea, L., Mataragas, M., & Drosinos, E. H. (2022). Effect of dough-related parameters on the antimold activity of *Wickerhamomyces anomalus* strains and mold-free shelf life of bread. *Applied Sciences, 12*(9), 4506. https://doi.org/10.3390/app12094506

Tan, H. P., Wan, T. S., Min, C. L. S., Osborne, M., & Ng, K. H. (2014). Quantitative analysis of fragrance in selectable one dimensional or two dimensional gas chromatography-mass spectrometry with simultaneous detection of multiple detectors in single injection. *Journal of Chromatography. A*, *1333*, 106–115. https://doi.org/10.1016/j.chroma.2014.01.073

Tao, N., Chen, Y., Wu, Y., Wang, X., Li, L., & Zhu, A. (2019). The terpene limonene induced the green mold of citrus fruit through regulation of reactive oxygen species (ROS) homeostasis in *Penicillium digitatum* spores. *Food Chemistry, 277*, 414–422. https://doi.org/10.1016/j.foodchem.2018.10.142

Tranchida, P. Q., Purcaro, G., Dugo, P., Mondello, L., & Purcaro, G. (2011). Modulators for comprehensive two-dimensional gas chromatography. *TrAC Trends in Analytical Chemistry, 30*(10), 1437–1461. https://doi.org/10.1016/j.trac.2011.06.010

Tranchida, P. Q., Salivo, S., Franchina, F. A., & Mondello, L. (2015). Flow-Modulated Comprehensive Two-Dimensional Gas Chromatography Combined with a High-Resolution Time-of-Flight Mass Spectrometer: A Proof-of-Principle Study. *Analytical Chemistry*, *87*(5), 2925–2930. https://doi.org/10.1021/ac5044175

Trinklein, T. J., & Synovec, R. E. (2022). Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis. *Journal of Chromatography A, 1677*, 463321. https://doi.org/10.1016/j.chroma.2022.463321

Trinklein, T. J., Cain, C. N., Ochoa, G. S., Schöneich, S., Mikaliunaite, L., & Synovec, R. E. (2023). Recent Advances in GC×GC and Chemometrics to Address Emerging Challenges in Nontargeted Analysis. *Analytical Chemistry*, *95*(1), 264–286. https://doi.org/10.1021/acs.analchem.2c04235

Trinklein, T. J., Gough, D. V., Warren, C. G., Ochoa, G. S., & Synovec, R. E. (2020). Dynamic pressure gradient modulation for comprehensive two-dimensional gas chromatography. *Journal of Chromatography A*, *1609*, 460488. https://doi.org/10.1016/j.chroma.2019.460488

Trinklein, T. J., Jiang, J., & Synovec, R. E. (2022). Profiling olefins in gasoline by bromination using GC×GC-TOFMS followed by discovery-based comparative analysis. *Analytical Chemistry, 94*(21), 9407–9414. https://doi.org/10.1021/acs.analchem.2c01549

Trinklein, T. J., Schöneich, S., Sudol, P. E., Warren, C. G., Gough, D. V., & Synovec, R. E. (2020). Total-transfer comprehensive three-dimensional gas chromatography with time-of-flight mass spectrometry. *Journal of Chromatography A*, *1634*, 461654. https://doi.org/10.1016/j.chroma.2020.461654

Tu, B. P., Kudlicki, A., Rowicka, M., & McKnight, S. L. (2005). Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes. *Science*, *310*(5751), 1152–1158. https://doi.org/10.1126/science.1120499

Tu, B. P., Mohler, R. E., Liu, J. C., Dombek, K. M., Young, E. T., Synovec, R. E., & McKnight, S. L. (2007). Cyclic changes in metabolic state during the life of a yeast cell. *Proceedings of the National Academy of Sciences*, *104*(43), 16886–16891. https://doi.org/10.1073/pnas.0708365104

Tuo, S., Liu, C., Wang, C., Kong, B., Lu, H., Zhong, K., Li, Y., Liu, W., & Yu, J. (2024). Evaluation of Fourier deconvolution ion mobility spectrometer as high-performance gas chromatography detector for the analysis of plant extract flavors. *Journal of Chromatography A*, *1714*, 464560. https://doi.org/10.1016/j.chroma.2023.464560

Urbina, F., Zorn, K. M., Brunner, D., & Ekins, S. (2021). Comparing the Pfizer Central Nervous System Multiparameter Optimization Calculator and a BBB Machine Learning Model. *ACS Chemical Neuroscience*, *12*(12), 2247–2253. https://doi.org/10.1021/acschemneuro.1c00265

Vogt, M., Brüsseler, C., Ooyen, J. van, Bott, M., & Marienhagen, J. (2016). Production of 2-methyl-1-butanol and 3-methyl-1-butanol in engineered *Corynebacterium glutamicum*. *Metabolic Engineering*, *38*, 436–445. https://doi.org/10.1016/j.ymben.2016.10.007

Wang, C.-H., Huang, C.-Y., Yak, H.-K., Hsieh, H.-C., & Wang, J.-L. (2021). Identifying an unknown compound in flue gas of semiconductor industry – Forensics of a perfluorocarbon. *Chemosphere, 264*, 128504. https://doi.org/10.1016/j.chemosphere.2020.128504

Wang, S., Chen, H., & Sun, B. (2020). Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS). *Food Chemistry*, *315*, 126158. https://doi.org/10.1016/j.foodchem.2019.126158

Wang, S., Kind, T., Bremer, P. L., Tantillo, D. J., & Fiehn, O. (2022). Quantum Chemical Prediction of Electron Ionization Mass Spectra of Trimethylsilylated Metabolites. *Analytical Chemistry*, *94*(3), 1559–1566. https://doi.org/10.1021/acs.analchem.1c02838

Want, E. J., Nordström, A., Morita, H., & Siuzdak, G. (2007). From Exogenous to Endogenous: The Inevitable Imprint of Mass Spectrometry in Metabolomics. *Journal of Proteome Research*, *6*(2), 459–468. https://doi.org/10.1021/pr060505+

Watanabe, R., Sugahara, A., Hagihara, H., Mizukado, J., & Shinzawa, H. (2021). Three-way evolved gas analysis-mass spectrometry combined with principal component analysis (EGA-MS-PCA) to probe interfacial states between matrix and filler in poly(styrene-b-butadiene-b-styrene) (SBS) nanocomposites. *Polymer Testing*, *101*, 107300. https://doi.org/10.1016/j.polymertesting.2021.107300

Welke, J. E., Nicolli, K. P., Hernandes, K. C., Biasoto, A. C. T., & Zini, C. A. (2022). Adaptation of an olfactometric system in a GC-FID in combination with GCxGC/MS to evaluate odor-active compounds of wine. *Food Chemistry*, *370*. https://doi.org/10.1016/j.foodchem.2021.131004

West, C. (2016). Statistics for analysts who hate statistics, Part III: Principal component analysis. *Chromatography Online, 34*(12), 868–869. https://www.chromatographyonline.com/view/statistics-analysts-who-hate-statistics-part-iii-principal-component-analysis

Yakovleva, E. Y., & Patrushev, Y. V. (2021). Determination of hydrocarbons in n-butane on porous layer capillary columns with nonpolar stationary phase. *Chromatographia, 84*(12), 1095–1104. https://doi.org/10.1007/s10337-021-04092-1

Yan, Y., Zhu, F., Zhu, C., Chen, Z., Liu, S., Wang, C., & Gu, C. (2021). Dibutyl phthalate release from polyvinyl chloride microplastics: Influence of plastic properties and environmental factors. *Water Research, 204*, 117597. https://doi.org/10.1016/j.watres.2021.117597

Yang, J., Park, S., Kim, H. J., Lee, S. J., & Jung, W. H. (2023). *The Interkingdom Interaction with Staphylococcus Influences the Antifungal Susceptibility of the Cutaneous Fungus Malassezia*. *33*(2), 180–187. https://doi.org/10.4014/jmb.2210.10039

Yost, E. E., Euling, S. Y., Weaver, J. A., Beverly, B. E. J., Keshava, N., Mudipalli, A., Arzuaga, X., Blessinger, T., Dishaw, L., Hotchkiss, A., & Makris, S. L. (2019). Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies. *Environmental International, 125*, 579–594. https://doi.org/10.1016/j.envint.2018.09.038

Yun, H., & Lee, M. L. (1996). Fast gas chromatography of light hydrocarbons and permanent gases on porous-layer open-tubular columns. *Field Analytical Chemistry & Technology, 1*(1), 60–64. https://doi.org/10.1002/(SICI)1520-6521(1996)1:1<60::AID-FACT8>3.0.CO;2-I

Yun, H., Lee, M. L., & Markides, K. E. (1995). Charcoal porous layer open tubular column gas chromatography for permanent gas analysis. *Journal of Microcolumn Separations*, *7*(3), 207–212. https://doi.org/10.1002/mcs.1220070304

Zanella, D., Focant, J., & Franchina, F. A. (2021). 30th Anniversary of comprehensive two-dimensional gas chromatography: Latest advances. *Analytical Science Advances*, *2*, 213–224. https://doi.org/doi.org/10.1002/ansa.202000142

Zarr. (n.d.). INSIGHT-Thermal modulator of GCxGC. Retrieved September 3, 2024, from https://www.sepsolve.com/gas-chromatography/insight-thermal-modulator-of-gcxgc.aspx

Zhang, Y., Lane, S., Chen, J.-M., Hammer, S. K., Luttinger, J., Yang, L., Jin, Y.-S., & Avalos, J. L. (2019). Xylose utilization stimulates mitochondrial production of isobutanol and 2-methyl-1-butanol in Saccharomyces cerevisiae. *Biotechnology for Biofuels*, *12*(1), 223. https://doi.org/10.1186/s13068-019-1560-2

Zhang, Y., Ren, H., Tang, X., Liu, Q., Xiao, W., Zhang, Z., & Tian, Y. (2024). A GC×GC-MS method based on solid-state modulator for non-targeted metabolomics: Comparison with traditional GC-MS method. *Journal of Pharmaceutical and Biomedical Analysis, 243*, 116068. https://doi.org/10.1016/j.jpba.2024.116068

Zhou, W., Sun, W., Liu, Y., Mao, Z., & Chen, Z. (2020). Ionic liquid-copolymerized monolith based porous layer open tubular column for CEC-MS analysis. *Talanta*, *209*, 120556. https://doi.org/10.1016/j.talanta.2019.120556

Zou, Y., Gaida, M., Franchina, F. A., Stefanuto, P. H., & Focant, J. F. (2022). Distinguishing between Decaffeinated and Regular Coffee by HS-SPME-GC×GC-TOFMS, Chemometrics, and Machine Learning. *Molecules*, *27*, 1806. https://doi.org/10.3390/molecules27061806

Zou, Y., Stefanuto, P.-H., Maimone, M., Janssen, M., & Focant, J.-F. (2021). Unraveling the complex olefin isomer mixture using two-dimensional gas chromatography-photoionization-time of flight mass spectrometry. *Journal of Chromatography A, 1645*, 462103. https://doi.org/10.1016/j.chroma.2021.462103

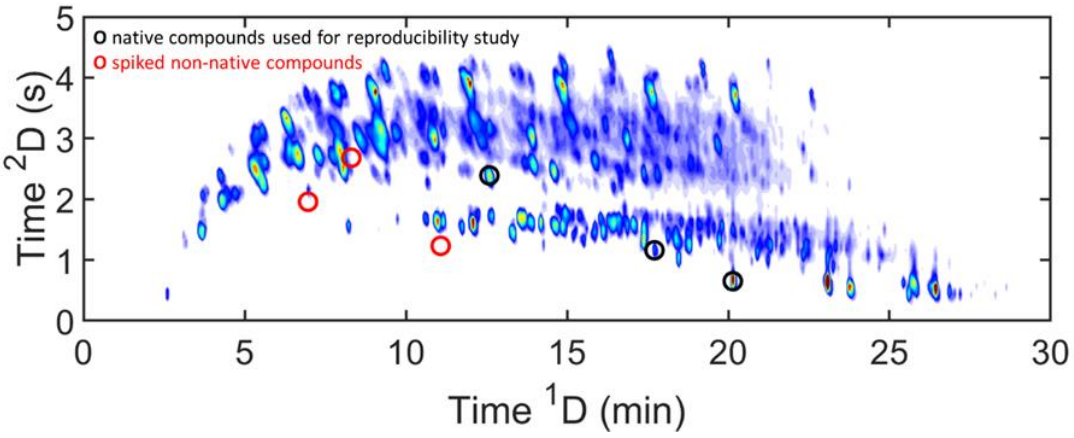
Appendix A

This appendix is reproduced from the Electronic Supplementary Material of L. Mikaliunaite, T. J. Trinklein, G. S. Ochoa, P. E. Sudol, D. S. Bell, R. E. Synovec, "Valve-based comprehensive two-dimensional gas chromatography with quadrupole mass spectrometry detection using a porous layer open tubular column in the first dimension" *J. Chromatogr. O* 3 (2023) 100076.

Table A.1. Summary of the experimental parameters used in peak reproducibility and LOD study, where JP8 fuel was analyzed using WCOT columns on both dimensions.

^{1}D Column	30 m Rxi-17SilMS, 250 µm d_c × 0.25 µm d_f
^{2}D Column	1.8 m Rtx-1, 100 µm d_c × 0.1 µm d_f
Sample	JP8 jet fuel
Injection volume	0.1 µL
Split ratio	No split
P_M (s) p_w (ms)	5 s 250 ms
Oven program	40 °C held for 1.5 min, then ramped to 200 °C at 5 °C/min and held for 1.5 min
Inlet pressure program, P_{inlet}	45.37 kPa (6.58 psig) ramped to 79.77 kPa (11.57 psig) at 1.10 kPa/min (0.24 psig/min)
Auxiliary pressure program, P_{aux}	26.89 kPa (3.9 psig) ramped to 80.67 kPa (11.70 psig) at 1.65 kPa/min (0.24 psig/min)

Figure A.1. Sample chromatogram of JP8 fuel indicating the location of three native analytes (black dots) used for peak reproducibility study (Fig. 3.9.) and location of three spiked non-native compounds (red dots) used for LOD study (Fig. 3.10.).



Appendix B

This appendix is reproduced from the Electronic Supplementary Material of L. Mikaliunaite, R. E. Synovec, "Computational method for untargeted determination of cycling yeast metabolites using comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry" *Talanta* 244 (2022) 123396.

Table B.1. Summary of the parameters used in the implementation of ChromaTOF Tile Software version 101 (LECO, St. Joseph, MI, USA) to perform tile-based F-ratio analysis, and quantify each metabolite hit in the hit list at each time-point.

ChromaTOF Tile method	
Tile size ^{1}D	10
Tile size ^{2}D	15
S/N threshold	10
Samples that must exceed *S/N* threshold	3
Mass F-ratios to average	1
Threshold type to apply	F-ratio
F-ratio threshold	0
Minimum masses per tile	1
Minimum mass	0
Maximum mass	10000

Figure B.1. Averaging 72 peak areas to 24 time-point signal sequences. Using the top F-ratio m/z for each of the 672 hits, the three replicates at each time-point were averaged to reduce the data density from 72 points in (A) to the final 24-point signal sequences in (B) for each of the hits (myo-inositol shown as an example). The averaged 24-point signal sequence is then used in all subsequent data analysis steps.

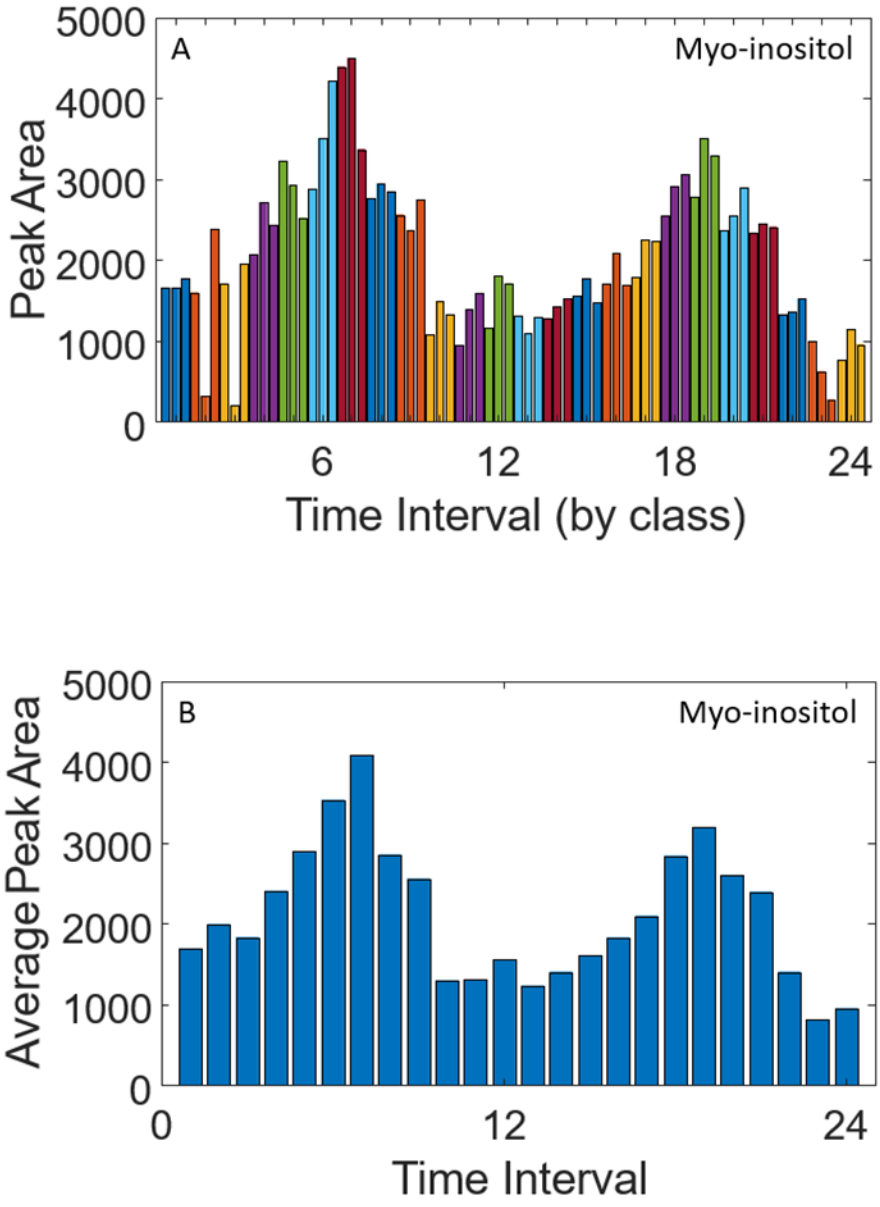


Figure B.2. To provide a robust and accurate determination of the LOF cycling pattern threshold, randomly generated signal sequences span a range of RSD to match the range of the experimental GC×GC-TOFMS data (both normalized to have a mean of 1). Here myo-inositol is used as an example to show that normalization of the data to the mean has no impact over LOF, where in (A) and (B) the LOF is calculated for non-normalized data, for comparison to (C) and (D), where the normalized peak areas are used. In (B) and (D), the LOF calculation is illustrated for the LOF determination for myo-inositol where the first 12 time-points (1 – 12) and the last 12 time-points (13 – 24) from (A) or (C) accordingly are matched up, 1 to 13, 2 to 14, etc., and then Eq. (1) is applied to result in the same LOF of 22.0%.

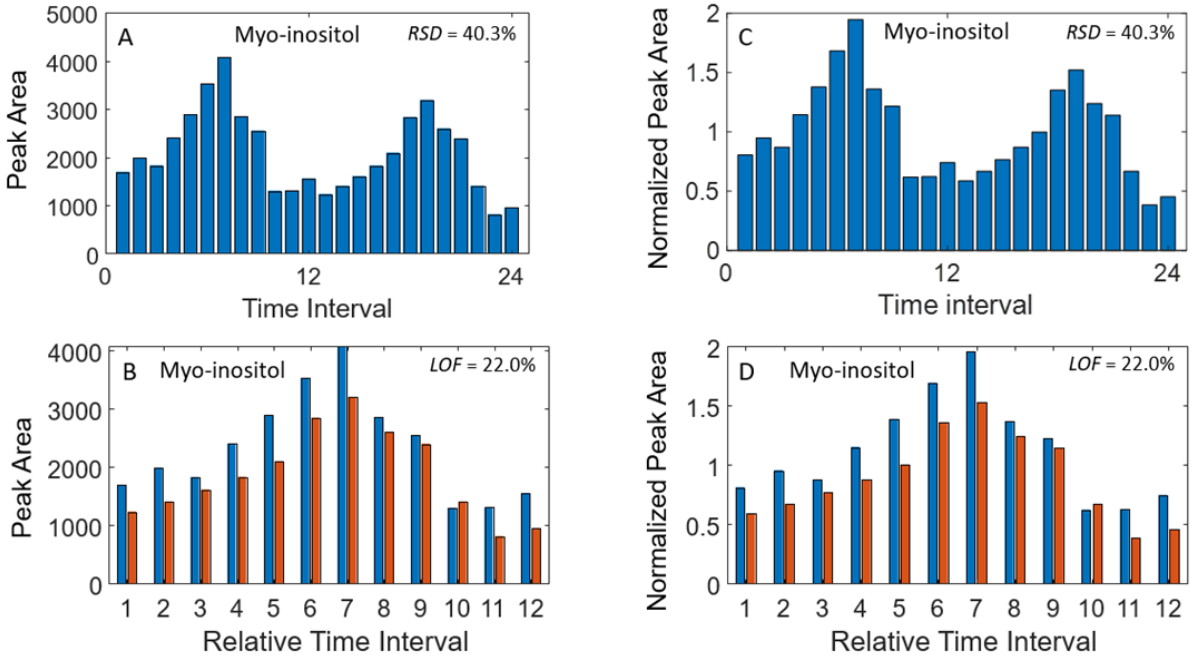


Figure B.3. LOF versus RSD statistical threshold determination was applied from a simulated RSD range of 10% to 100%, the resulting 95% confidence interval for LOF were fit to an equation and extrapolated to slightly higher RSD using Matlab 2019b using Curve Fitting Toolbox. The polynomial model, as well as goodness-of-fit parameters are also provided.

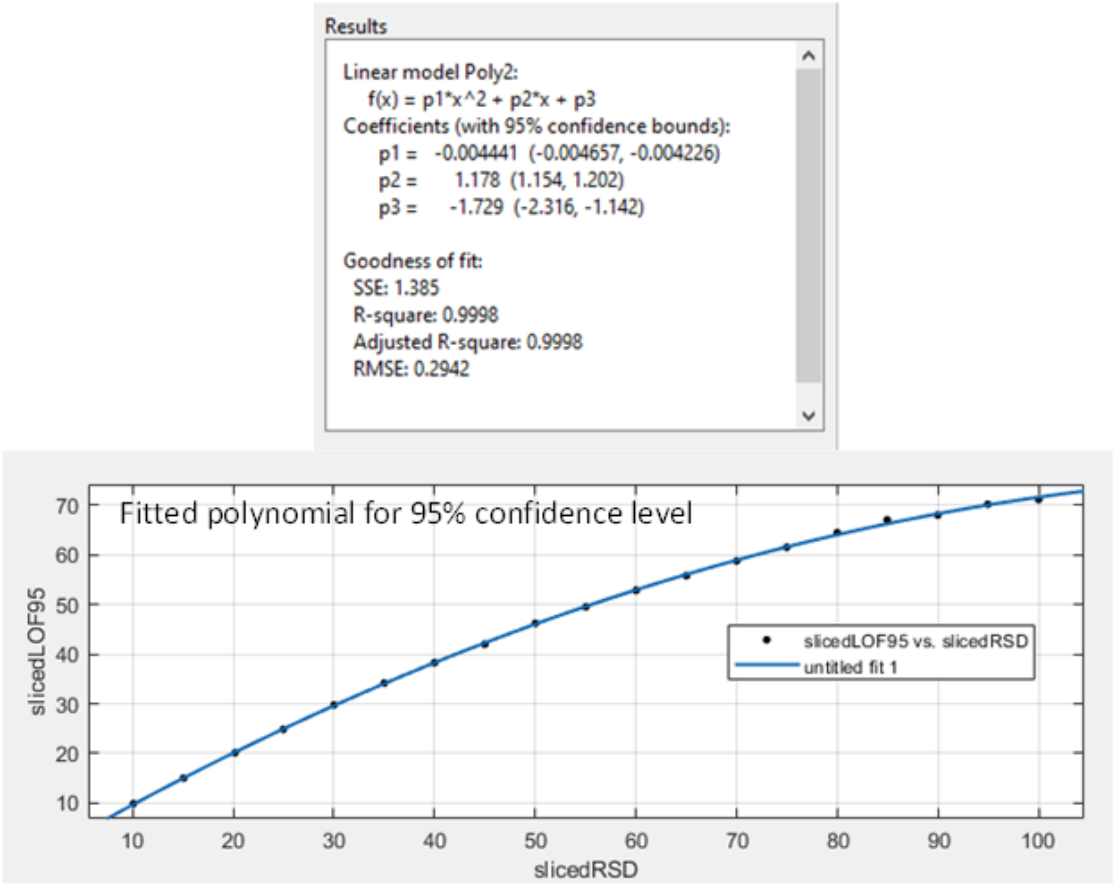


Table B.2. Final hitlist for the 155 unknown metabolites, where metabolites are ordered by category: 1 cycle/period (1c), 2 cycle/period (2c), spiky (sp) and multimodal (mm) and within each category is ordered by increasing phase (degrees).

Phase	Name	Phase	Name	Phase	Name	Phase	Name	Phase	Name
1c, 0	Unk 1	1c, 150	Unk 32	1c, 210	Unk 63	mm	Unk 94	mm	Unk 125
1c, 0	Unk 2	1c, 150	Unk 33	1c, 210	Unk 64	mm	Unk 95	mm	Unk 126
1c, 0	Unk 3	1c, 150	Unk 34	1c, 240	Unk 65	mm	Unk 96	mm	Unk 127
1c, 30	Unk 4	1c, 180	Unk 35	1c, 240	Unk 66	mm	Unk 97	mm	Unk 128
1c, 30	Unk 5	1c, 180	Unk 36	1c, 270	Unk 67	mm	Unk 98	mm	Unk 129
1c, 30	Unk 6	1c, 180	Unk 37	1c, 270	Unk 68	mm	Unk 99	mm	Unk 130
1c, 30	Unk 7	1c, 180	Unk 38	1c, 270	Unk 69	mm	Unk 100	mm	Unk 131
1c, 30	Unk 8	1c, 180	Unk 39	1c, 270	Unk 70	mm	Unk 101	mm	Unk 132
1c, 60	Unk 9	1c, 180	Unk 40	1c, 270	Unk 71	mm	Unk 102	mm	Unk 133
1c, 60	Unk 10	1c, 180	Unk 41	2c, 90	Unk 72	mm	Unk 103	mm	Unk 134
1c, 60	Unk 11	1c, 180	Unk 42	2c, 90	Unk 73	mm	Unk 104	mm	Unk 135
1c, 60	Unk 12	1c, 180	Unk 43	2c, 90	Unk 74	mm	Unk 105	mm	Unk 136
1c, 60	Unk 13	1c, 180	Unk 44	2c, 120	Unk 75	mm	Unk 106	mm	Unk 137
1c, 60	Unk 14	1c, 180	Unk 45	2c, 150	Unk 76	mm	Unk 107	mm	Unk 138
1c, 60	Unk 15	1c, 180	Unk 46	2c, 180	Unk 77	mm	Unk 108	mm	Unk 139
1c, 60	Unk 16	1c, 180	Unk 47	2c, 180	Unk 78	mm	Unk 109	mm	Unk 140
1c, 60	Unk 17	1c, 180	Unk 48	2c, 180	Unk 79	mm	Unk 110	mm	Unk 141
1c, 90	Unk 18	1c, 180	Unk 49	2c, 240	Unk 80	mm	Unk 111	mm	Unk 142
1c, 90	Unk 19	1c, 180	Unk 50	2c, 240	Unk 81	mm	Unk 112	mm	Unk 143
1c, 90	Unk 20	1c, 180	Unk 51	2c, 240	Unk 82	mm	Unk 113	mm	Unk 144
1c, 90	Unk 21	1c, 180	Unk 52	2c, 240	Unk 83	mm	Unk 114	mm	Unk 145
1c, 120	Unk 22	1c, 180	Unk 53	sp,120	Unk 84	mm	Unk 115	mm	Unk 146
1c, 120	Unk 23	1c, 210	Unk 54	sp, 120	Unk 85	mm	Unk 116	mm	Unk 147
1c, 120	Unk 24	1c, 210	Unk 55	sp, 270	Unk 86	mm	Unk 117	mm	Unk 148
1c, 120	Unk 25	1c, 210	Unk 56	sp, mm	Unk 87	mm	Unk 118	mm	Unk 149
1c, 120	Unk 26	1c, 210	Unk 57	sp, mm	Unk 88	mm	Unk 119	mm	Unk 150
1c, 120	Unk 27	1c, 210	Unk 58	sp, mm	Unk 89	mm	Unk 120	mm	Unk 151
1c, 120	Unk 28	1c, 210	Unk 59	sp, mm	Unk 90	mm	Unk 121	mm	Unk 152
1c, 120	Unk 29	1c, 210	Unk 60	mm	Unk 91	mm	Unk 122	mm	Unk 153
1c, 150	Unk 30	1c, 210	Unk 61	mm	Unk 92	mm	Unk 123	mm	Unk 154
1c, 150	Unk 31	1c, 210	Unk 62	mm	Unk 93	mm	Unk 124	mm	Unk 155

Figure B.4. Heatmap of the 155 unknown metabolites ordered according to Table B.2., where the signal for each time interval is shown where white lines are separating different types of metabolites; lowest concentration is in blue and highest concentration is in red.

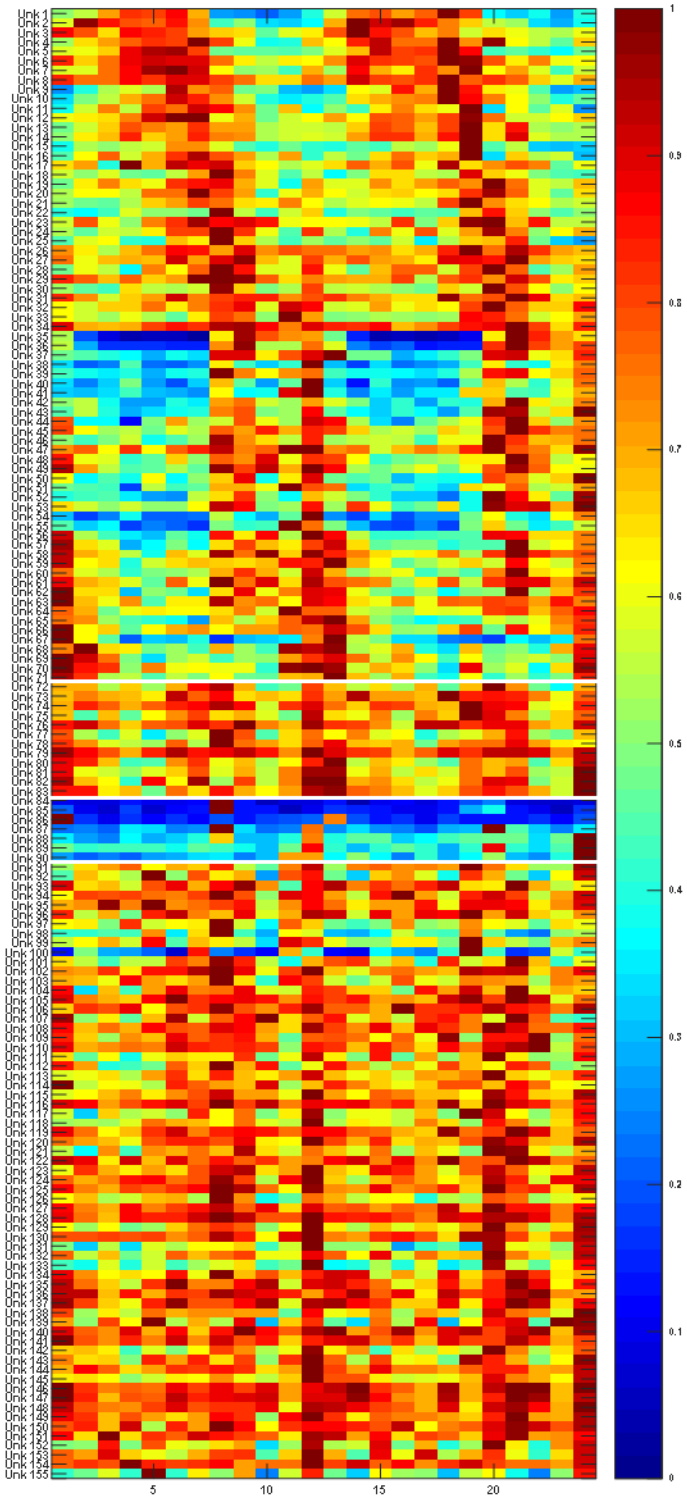
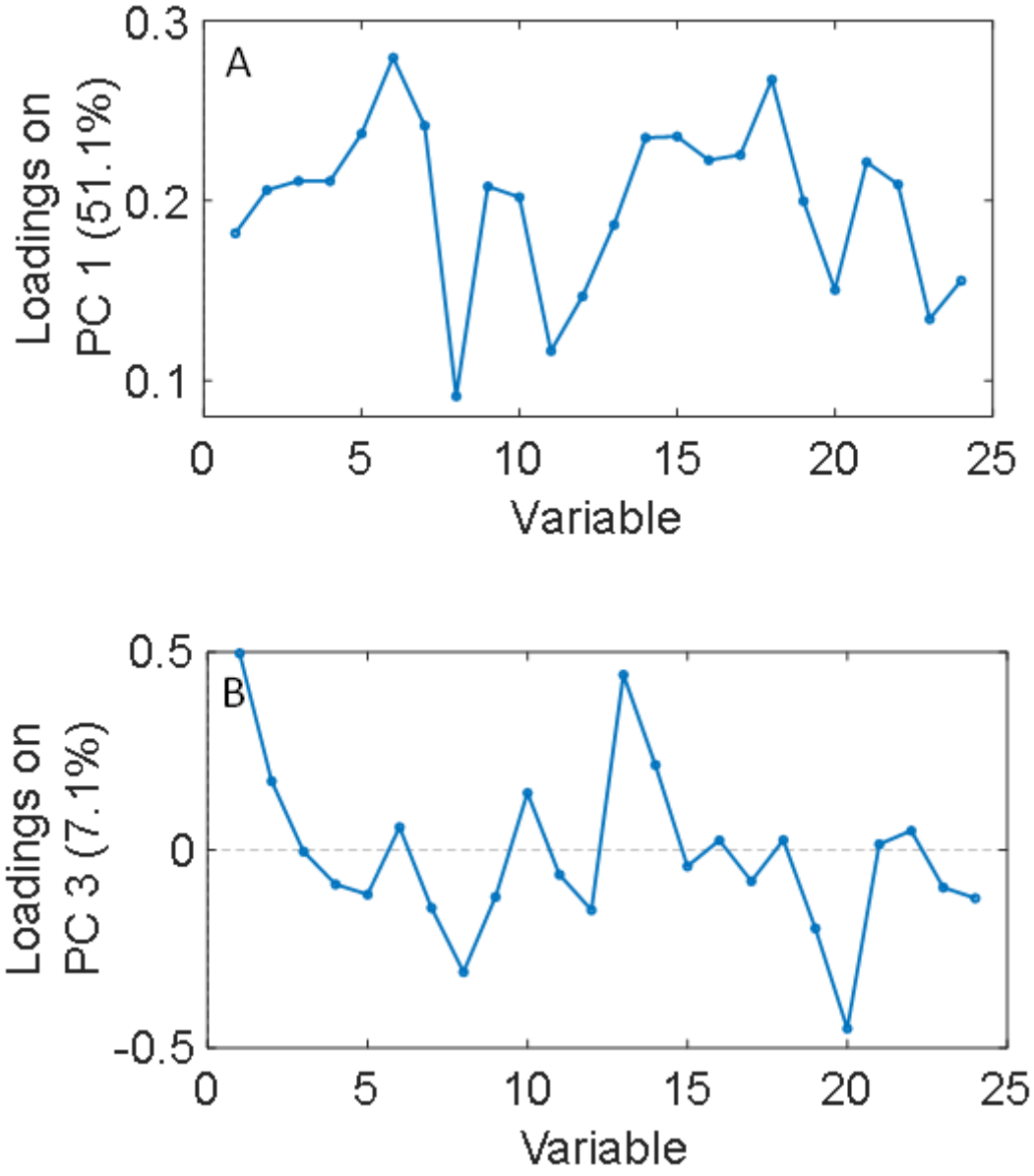


Figure B.5. PC 1 (A) and PC3 (B) loadings, showing a somewhat repeating cycling pattern between the first 12 time-points and the last 12 time-points.



Appendix C

Figure C.1. TIC chromatograms of blank at (A) pH 5.7, (B) pH 9.7, (C) pH 12.4, and of M. pachydermatis *at (D) pH 5.7, (E) pH 9.7, (F) pH 12.4.*

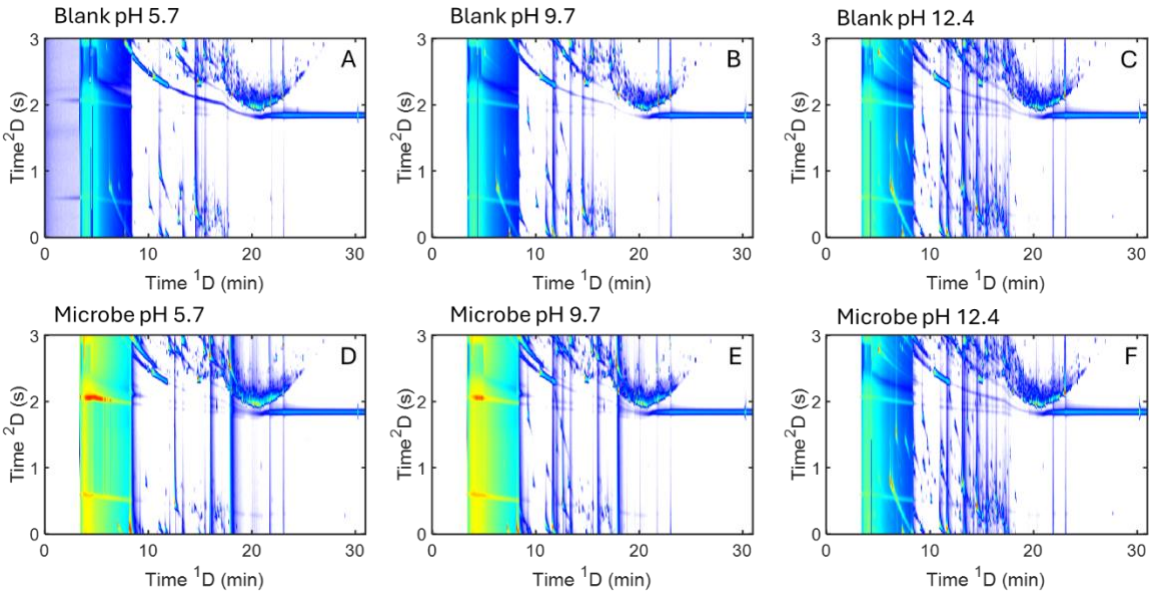


Figure C.2. Zoom-in of TIC chromatograms of blank at (A) pH 5.7, (B) pH 9.7, (C) pH 12.4, and of M. pachydermatis at (D) pH 5.7, (E) pH 9.7, (F) pH 12.4.

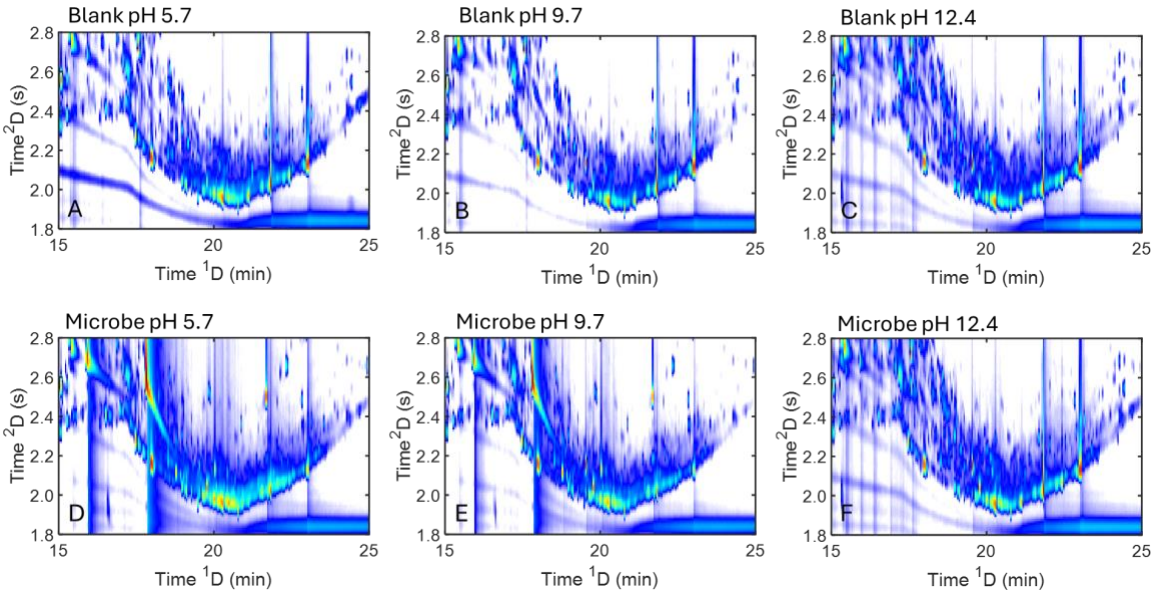


Table C.1. Standard parameters used to build for support vector machine regression model.

Parameters	Values
Gamma	0.000001, 0.00000316228, 0.00001, 0.0000316228, 0.0001, 0.000316228, 0.001, 0.00316228, 0.01, 0.0316228, 0.1, 0.316228, 1, 3.16228, 10
Cost	0.001, 0.003162278, 0.01, 0.03162278, 0.1, 0.3162278, 1, 3.162278, 100
Epsilon	1, 0.1, 0.01
nu	0.2, 0.5, 0.8

Table C.2. The final hitlist of 566 analytes, with their identification and Fisher ratio (F-ratio).

Hit #	Compound	CAS-number	F-ratio
1	Phenylethyl Alcohol	60-12-8	2538.4
2	Furan, 2-methyl-	534-22-5	1423.9
3	Unknown		1406.5
4	Furan, 2,5-dimethyl-	625-86-5	1299.1
5	Ethanone, 1-(1H-pyrazol-4-yl)-	25015-16-4	1237.0
6	2-Furanmethanol	98-00-0	1035.0
7	2-Furanmethanol, 5-methyl-	3857-25-8	910.4
8	Unknown		860.6
9	Unknown		755.7
10	Unknown		738.8
11	Unknown		718.7
12	Unknown		594.4
13	Unknown		556.3
14	Unknown		519.1
15	2-Naphthalenol	135-19-3	500.0
16	Unknown		474.2
17	Unknown		473.5
18	Unknown		459.8
19	Unknown		457.9
20	Unknown		430.9
21	Unknown		406.1
22	2-Butanone	78-93-3	395.5
23	Pyrazine	290-37-9	384.3
24	Unknown		379.8
25	Methallyl cyanide	4786-19-0	370.4
26	Pyrazine, methyl-	109-08-0	354.2
27	Unknown		348.5
28	Unknown		345.1
29	Hexanal	66-25-1	324.5
30	Unknown		307.5
31	2,4-Heptanedione	7307-02-0	299.1
32	2,5-Hexanedione	110-13-4	298.1
33	Benzofuran, 7-methyl-	17059-52-8	290.1
34	2-Cyclopenten-1-one, 3,5,5-trimethyl-	24156-95-4	289.3
35	Unknown		281.1
36	1,3-Diazine	289-95-2	279.4
37	2-Cyclopenten-1-one, 3-ethyl-2-hydroxy-	21835-01-8	278.4
38	4-Methylpyridazine	1120-88-3	276.3
39	Unknown		263.9
40	Unknown		258.8
41	p-Cymen-7-ol	536-60-7	252.1
42	Furan, 2,3,5-trimethyl-	10504-04-8	251.2
43	1,2,4,4-Tetramethylcyclopentene	65378-76-9	249.6
44	Dothiepin sulfoxide	1447-71-8	249.4
45	Unknown		243.6
46	3-Methylene-2-norbornanone	5597-27-3	243.0
47	Unknown		242.0
48	Benzene, 1-methoxy-3-methyl-	100-84-5	241.0
49	N-Dimethylaminomethyl-tert.-butyl-isopropylphosphine	83718-54-1	232.4
50	Unknown		232.4

51	Unknown		232.0
52	Unknown		230.3
53	2-Cyclopenten-1-one, 3,4,4-trimethyl-	30434-65-2	225.9
54	Unknown		224.9
55	(3Z,5E)-1,3,5-Undecatriene	19883-27-3	218.6
56	Unknown		217.6
57	2-Cyclopenten-1-one	930-30-3	217.2
58	Unknown		215.8
59	1-Butanol, 3-methyl-	123-51-3	215.4
60	Unknown		215.2
61	Ethanol	64-17-5	214.5
62	Cyclopropane, 1,2-dimethyl-, trans-	2402-06-4	213.5
63	Acetoin	513-86-0	209.2
64	Unknown		206.4
65	Unknown		205.3
66	1,2-Cyclopentanedione, 3-methyl-	765-70-8	203.8
67	Unknown		201.5
68	Ethanone, 1-(1-cyclohexen-1-yl)-	932-66-1	199.1
69	Formic acid, 2-phenylethyl ester	104-62-1	199.0
70	Unknown		196.7
71	2-Cyclopenten-1-one, 2-methyl-	1120-73-6	196.6
72	Ethanone, 1-(2-methyl-1-cyclopenten-1-yl)-	3168-90-9	195.9
73	Unknown		193.6
74	Unknown		193.5
75	Cyclopentanone, 3,4-bis(methylene)-	27646-73-7	191.8
76	Unknown		191.6
77	1,2-Difluoroethane	624-72-6	190.1
78	Unknown		185.2
79	Benzaldehyde	100-52-7	179.0
80	Propane, 1-(methylthio)-	3877-15-4	176.5
81	2-Hydroxy-3-pentanone	5704-20-1	175.3
82	Unknown		173.3
83	Unknown		169.8
84	3-Furaldehyde	498-60-2	166.9
85	Unknown		165.7
86	Unknown		164.9
87	Unknown		162.3
88	1-Pentanol	71-41-0	161.3
89	Unknown		160.6
90	3,4-Pentadienal	4009-55-6	160.5
91	1-Butanol, 2-methyl-, (S)-	1565-80-6	160.0
92	Furan, 2-ethyl-5-methyl-	1703-52-2	159.8
93	Unknown		158.5
94	2-Cyclopenten-1-one, 3,4,5-trimethyl-	55683-21-1	157.5
95	Unknown 03745		155.8
96	Bicyclo[4.2.0]octa-1,3,5-triene	694-87-1	154.4
97	Unknown		152.8
98	Unknown		151.7
99	Oxirane, 2-(1,1-dimethylethyl)-3-methyl-	53897-30-6	148.1
100	1-Hydroxy-2-butanone	5077-67-8	147.7
101	Unknown		147.1
102	Unknown		143.3
103	Unknown		143.1

104	2-Acetylthiazole	24295-03-2	142.4
105	3-Pentanol	584-02-1	140.9
106	Cyclohexene, 4-methyl-1-(1-methylethyl)-	500-00-5	140.2
107	Unknown		138.7
108	Unknown		136.9
109	Nonadecane	629-92-5	136.8
110	Unknown		136.7
111	Unknown		136.1
112	1-Pentanol, 2-ethyl-4-methyl-	106-67-2	135.5
113	Butane, 2,2,3,3-tetramethyl-	594-82-1	135.4
114	Unknown		134.6
115	Unknown		132.5
116	Dodecane, 2,6,10-trimethyl-	3891-98-3	132.5
117	7,9-Di-tert-butyl-1-oxaspiro(4,5)deca-6,9-diene-2,8-dione	82304-66-3	132.3
118	Unknown		131.8
119	Methanethiol	74-93-1	130.4
120	2-Cyclohexen-1-one, 3-methyl-	1193-18-6	129.5
121	Unknown		128.1
122	Furfural	98-01-1	128.1
123	Cyclopentane, 1-acetyl-1,2-epoxy-	15121-02-5	126.5
124	Unknown		125.6
125	Unknown		124.6
126	Pyrazole-4-carboxaldehyde, 1-methyl-	25016-11-9	124.3
127	Unknown		123.2
128	1,3-Cyclopentanedione, 2,4-dimethyl-	34598-80-6	122.7
129	Butane, 2-cyclopropyl-	5750-02-7	121.6
130	Unknown		121.4
131	Furan, 2-(2-propenyl)-	75135-41-0	121.1
132	Unknown		120.0
133	2-Hexanone	591-78-6	120.0
134	2,3-Octanedione	585-25-1	119.8
135	Unknown		118.7
136	Unknown		118.4
137	2-Octene	111-67-1	118.3
138	Unknown		118.0
139	1-Adamantanol	768-95-6	117.8
140	Benzaldehyde, 2-hydroxy-	90-02-8	115.4
141	Unknown		114.4
142	1H-Pyrazole, 1,3-dimethyl-	694-48-4	113.8
143	Furan, 3-pentyl-	6177-84-0	113.7
144	Unknown		112.2
145	Nonadecane	629-92-5	109.4
146	2,4-Hexadiene, 2,5-dimethyl-	764-13-6	109.3
147	Unknown		109.0
148	1-Hexanol, 2-ethyl-	104-76-7	108.9
149	Unknown		108.3
150	Unknown		108.2
151	Unknown		108.1
152	2-Cyclopenten-1-one, 3,4-dimethyl-	30434-64-1	105.5
153	3-Methylpyridazine	1632-76-4	105.2
154	Unknown		104.7
155	3-Hexen-2-one	763-93-9	103.8
156	Unknown		103.3

157	trans-2-(2-Pentenyl)furan	70424-14-5	103.2
158	Unknown		103.2
159	2-Octen-4-ol, (E)-	20125-81-9	103.0
160	Benzyl alcohol	100-51-6	102.8
161	Unknown		102.6
162	Octane, 3,5-dimethyl-	15869-93-9	101.8
163	1H-Pyrazole, 3-methyl-	1453-58-3	101.7
164	Unknown		100.8
165	Unknown		100.2
166	1H-Pyrazole, 1-methyl-	930-36-9	100.2
167	Pyrazole, 1,4-dimethyl-	1072-68-0	100.1
168	Ethanone, 1-(2-furanyl)-	1192-62-7	100.0
169	Decane	124-18-5	97.2
170	Unknown		97.2
171	Unknown		97.2
172	1,2-Ethanediamine, N'-ethyl-N,N-dimethyl-	123-83-1	96.9
173	Furan, tetrahydro-2,5-dimethyl-, cis-	2144-41-4	96.6
174	2-Propanone, 1-hydroxy-	116-09-6	95.5
175	Unknown		95.4
176	1-Cyclohexene-1-methanol	4845-04-9	95.2
177	2-Cyclopenten-1-one, 3-methyl-	2758-18-1	93.5
178	Pyridine, 4-methyl-	108-89-4	93.4
179	2-Heptanone	110-43-0	93.1
180	Unknown		92.8
181	Unknown		92.1
182	Unknown		91.8
183	Unknown		91.4
184	Bicyclo[2.2.2]octan-1-ol, 4-methyl-	824-13-5	90.8
185	2-Hexanone, 4-methyl-	105-42-0	90.1
186	Unknown		89.5
187	Unknown		89.2
188	Unknown		88.5
189	2,4-Decadienal	2363-88-4	88.1
190	Unknown		88.1
191	Cyclobutene, 2-propenylidene-	52097-85-5	88.1
192	Unknown		88.0
193	Unknown		87.9
194	Furan, 2-ethyl-	3208-16-0	87.3
195	Unknown		87.1
196	Unknown		86.8
197	2(3H)-Furanone, 5-ethyldihydro-	695-06-7	86.6
198	Phenol, 3,5-dimethyl-	108-68-9	86.1
199	Hept-3-yn-2-one	26059-43-8	85.8
200	Unknown		85.7
201	Unknown		85.6
202	Furan, 2-pentyl-	3777-69-3	84.8
203	Unknown		84.8
204	Phenol, 3-methyl-	108-39-4	83.9
205	Cyclohexanol, 1-(2-hexenyl)-	54655-47-9	83.8
206	3-Octene, (Z)-	14850-22-7	83.8
207	Unknown		83.0
208	Dihydro-3-(2H)-thiophenone	1003-04-9	82.7
209	Unknown		82.6

210	Unknown		81.7
211	Unknown		81.5
212	Unknown		81.4
213	Heptanal	111-71-7	81.4
214	Unknown		81.2
215	Methylamine, N,N-dimethyl-	75-50-3	80.7
216	Unknown		80.7
217	7-Oxabicyclo[4.1.0]heptane, 2-methylene-	98126-49-9	80.6
218	2-Cyclopenten-1-one, 2,3-dimethyl-	1121-05-7	80.4
219	Unknown		80.4
220	Unknown		80.1
221	Unknown		79.6
222	Unknown		79.6
223	2,3-Dimethyldecane	17312-44-6	79.5
224	Unknown		79.0
225	2,4-Nonadienal, (E,E)-	5910-87-2	79.0
226	Benzene, 1,3-bis(1,1-dimethylethyl)-	1014-60-4	78.8
227	1,3-Oxathiane	646-12-8	78.5
228	Unknown		78.5
229	Pyrazine, 2,5-dimethyl-	123-32-0	78.5
230	Furan, 2,2'-methylenebis-	1197-40-6	78.0
231	Unknown		77.8
232	Acetylacetone-	123-54-6	77.3
233	Unknown		77.3
234	Unknown		77.0
235	2,3-Dimethyl-4-hydroxy-2-butenoic lactone	1575-46-8	76.8
236	Octane, 4-chloro-	999-07-5	76.7
237	1-Propanone, 1-(2-furanyl)-	3194-15-8	76.4
238	Unknown		76.4
239	1-Hexanol	111-27-3	76.4
240	Unknown		76.4
241	Furan, 3-(4-methyl-3-pentenyl)-	539-52-6	75.5
242	Propane, 2-bromo-1-chloro-	3017-95-6	75.4
243	1-Methylcyclopropanemethanol	2746-14-7	75.0
244	Unknown		74.6
245	Unknown		74.5
246	Unknown		74.4
247	trans-2-Oxabicyclo[4.4.0]decane	59958-46-2	74.2
248	2-Furancarboxaldehyde, 5-methyl-	620-02-0	73.9
249	Unknown		73.9
250	Nonanal	124-19-6	73.9
251	Unknown		73.3
252	Unknown		73.2
253	3-Hexen-2-one, 5-methyl-	5166-53-0	73.2
254	Unknown		73.0
255	Unknown		72.9
256	Cyclohexanone	108-94-1	72.5
257	3(2H)-Furanone, dihydro-2-methyl-	3188-00-9	72.5
258	Pyridine, 2-chloro-6-(2-furanylmethoxy)-4-(trichloromethyl)-	70166-48-2	72.4
259	Unknown		72.3
260	1,4-Hexadiene, 2-methyl-	1119-14-8	72.0
261	2-Octenal, (E)-	2548-87-0	71.8
262	Unknown		71.8

263	2-Methyl-1-octen-3-yne	17603-76-8	71.5
264	Unknown		71.3
265	3-Penten-2-one, 3-ethyl-4-methyl-	22287-11-2	71.2
266	Unknown		70.9
267	Unknown		70.9
268	Unknown		70.9
269	Unknown		70.6
270	Unknown		70.3
271	Unknown		70.2
272	Unknown		70.2
273	Bicyclo[3.1.0]hexane, 6-isopropylidene-	24524-58-1	70.1
274	Formic acid, butyl ester	592-84-7	70.1
275	Pentanal	110-62-3	69.9
276	Toluene	108-88-3	69.7
277	2-Cyclohexen-1-one	930-68-7	69.2
278	L-Leucine, methyl ester	2666-93-5	69.0
279	2(3H)-Furanone	20825-71-2	69.0
280	2-Heptenal, (Z)-	57266-86-1	68.9
281	Unknown		68.8
282	1,4-Hexadiene, 2-methyl-	1000-86-8	68.5
283	Unknown		68.2
284	3,6-Heptanedione	1703-51-1	68.1
285	Unknown		67.9
286	1-(2,4-Dimethyl-furan-3-yl)-ethanone	32933-07-6	67.6
287	Unknown		67.3
288	Resorcinol, 2-acetyl-	699-83-2	67.2
289	Unknown		67.0
290	Unknown		66.9
291	Unknown		66.5
292	Unknown		65.8
293	2-Methoxy-5-methylphenol	1195-09-1	65.6
294	Unknown		65.6
295	Hexadecane	544-76-3	65.5
296	2(3H)-Furanone, 5-methyl-	591-12-8	65.4
297	2-Hexanoylfuran	14360-50-0	65.3
298	Unknown		65.2
299	1-Propanol, 3-(methylthio)-	505-10-2	65.2
300	Cyclodecene, (Z)-	935-31-9	65.1
301	Unknown		64.8
302	Unknown		64.4
303	Unknown		63.8
304	Unknown		63.7
305	3-Acetyl-2,5-dimethyl furan	10599-70-9	63.4
306	1-Propanol, 2-amino-2-methyl-	124-68-5	63.3
307	1-Methylcyclooctene	933-11-9	63.3
308	Unknown		62.8
309	Unknown		62.2
310	Unknown		62.1
311	Unknown		62.0
312	Pyrimidine, 4-methyl-	3438-46-8	61.8
313	Unknown		61.6
314	Unknown		61.4
315	2H-Pyran, 3,4-dihydro-	110-87-2	61.3

316	3,5-Dimethylpyrazole	67-51-6	61.3
317	Unknown		61.2
318	1,3-Benzenediol, 4-ethyl-	2896-60-8	60.7
319	l9-Oxabicyclo[6.1.0]nonane	286-62-4	60.1
320	Unknown		59.9
321	Unknown		59.9
322	Unknown		59.9
323	Unknown		59.7
324	Pent-2-ynal	55136-52-2	59.6
325	1H-Indene-4-carboxaldehyde, 2,3-dihydro-	51932-70-8	59.6
326	5-Ethylcyclopent-1-enecarboxaldehyde	36431-60-4	59.5
327	Unknown		59.4
328	1,4-Benzenediol, 2,6-dimethyl-	654-42-2	59.1
329	Unknown		58.9
330	Unknown		58.5
331	Unknown		58.4
332	5-Hexen-2-one	109-49-9	58.3
333	1-Heptanol	111-70-6	58.3
334	3,5-Heptadien-2-one, 6-methyl-, (E)-	16647-04-4	58.0
335	Unknown		58.0
336	Hydrazine, (phenylmethyl)-	555-96-4	57.7
337	Unknown		57.6
338	o-Hydroxybiphenyl	90-43-7	57.5
339	Unknown		57.2
340	Unknown		56.8
341	1,4-Benzenediol, 2,5-dimethyl-	615-90-7	56.7
342	1H-Pyrazole, 1,5-dimethyl-	694-31-5	56.7
343	Unknown		56.4
344	Unknown		56.2
345	Unknown		56.2
346	Unknown		56.1
347	Unknown		56.0
348	Unknown		55.8
349	Benzene, butoxy-	1126-79-0	55.5
350	Cyclopropane, 1,2-dimethyl-, cis-	930-18-7	55.5
351	1,3-Dimethyl-1-cyclohexene	2808-76-6	55.4
352	Benzofuran, 2-methyl-	4265-25-2	55.0
353	2(5H)-Furanone	497-23-4	55.0
354	Unknown		55.0
355	Unknown		54.8
356	Pyrazine, ethyl-	13925-00-3	54.7
357	2(3H)-Furanone, 5-butyldihydro-	104-50-7	54.6
358	Octane	111-65-9	54.5
359	1-Propanone, 1-(5-methyl-2-furanyl)-	10599-69-6	54.4
360	Benzene, 1,4-dimethoxy-2-methyl-	24599-58-4	53.9
361	Unknown		53.5
362	Phenol	108-95-2	53.3
363	4-Cyclopentene-1,3-dione	930-60-9	53.2
364	Unknown		52.9
365	Cyclooctene	931-88-4	52.8
366	Unknown		52.7
367	Pyrimidine, 4,5-dimethyl-	694-81-5	52.6
368	1,4-Cyclohex-2-enedione	4505-38-8	52.6

369	2-Pentenal, (E)-	1576-87-0	52.5
370	Methylthio-2-propanone	14109-72-9	52.3
371	2-Propyl-1-pentanol	58175-57-8	52.3
372	Unknown		52.2
373	Unknown		52.1
374	Unknown		52.1
375	Unknown		52.0
376	Unknown		51.9
377	Unknown		51.9
378	Unknown		51.8
379	3-Octen-2-one	1669-44-9	51.7
380	Bicyclo[5.1.0]octane	286-43-1	51.5
381	Furan, 2-hexyl-	3777-70-6	51.4
382	Unknown		51.3
383	Unknown		51.3
384	Unknown		51.2
385	Oxepine, 2,7-dimethyl-	1487-99-6	51.0
386	Unknown		51.0
387	Oxirane, (1-methylbutyl)-	53229-39-3	51.0
388	Cyclohexene, 3-methyl-6-(1-methylethyl)-, trans-	1124-26-1	51.0
389	2(5H)-Furanone, 5-methyl-	591-11-7	51.0
390	Unknown		50.4
391	2-Butene, 2-methyl-	513-35-9	49.9
392	2-Furanone, 2,5-dihydro-3,5-dimethyl	5584-69-0	49.8
393	1-Hexanol, 4-methyl-	818-49-5	49.4
394	2(3H)-Furanone, 5-hexyldihydro-	706-14-9	49.2
395	Maleic anhydride	108-31-6	49.0
396	3-Cyclobutene-1,2-dione	32936-74-6	48.9
397	Unknown		48.9
398	2-Penten-1-ol, (E)-	1576-96-1	48.5
399	Unknown		48.5
400	Unknown		48.4
401	Benzoic acid, methyl ester	93-58-3	48.4
402	Unknown		48.3
403	Unknown		48.0
404	Formic acid, 2-methylbutyl ester	35073-27-9	48.0
405	Octanal	124-13-0	47.6
406	2-Butenal, 3-methyl-	107-86-8	47.3
407	2-Furoic acid, 2,3,4,6-tetrachlorophenyl ester	N/A	47.3
408	Unknown		47.2
409	1-ethoxy-2,4-hexadiene	56752-54-6	47.2
410	Unknown		47.1
411	Unknown		47.1
412	Unknown		47.0
413	Unknown		46.9
414	Undecane, 4-methyl-	2980-69-0	46.9
415	y-Dodecalactone	2305-05-7	46.7
416	Unknown		46.7
417	Unknown		46.4
418	Undecane, 2,6-dimethyl-	17301-23-4	46.1
419	Unknown		46.0
420	2,5-Cyclohexadiene-1,4-dione, 2,6-bis(1,1-dimethylethyl)-	719-22-2	46.0
421	Unknown		45.9

422	4-Methyl-5H-furan-2-one	6124-79-4	45.9
423	1,2,4-Triazin-3-amine, 5,6-dimethyl-	17584-12-2	45.6
424	Unknown		45.3
425	Isophorone	78-59-1	45.1
426	2-Furanmethanol, acetate	623-17-6	45.0
427	1,2-Benzenedicarboxaldehyde	643-79-8	45.0
428	1,3-Octadiene	1002-33-1	44.8
429	Unknown		44.5
430	Unknown		44.4
431	Unknown		44.3
432	Unknown		44.2
433	1-Hexanol, 5-methyl-	627-98-5	44.1
434	2,4-Hexadiene, 2,5-dimethyl-	1515-79-3	44.0
435	Unknown		44.0
436	Ethanone, 1-(3-thienyl)-	1468-83-3	43.9
437	1,3-Cyclopentadiene, 5-(1-methylethylidene)-	2175-91-9	43.7
438	2-Dodecanone	6175-49-1	43.6
439	Furan, 2,3-dihydro-5-methyl-	1487-15-6	43.3
440	Unknown		43.3
441	2(3H)-Furanone, dihydro-5-(2-octenyl)-, (Z)-	18679-18-0	42.5
442	Cyclopentene, 1-(1-methylethyl)-	1462-07-3	42.4
443	Unknown		42.2
444	Unknown		42.1
445	4-Hydroxy-2-butanone	590-90-9	42.0
446	Acetylpyrazine	22047-25-2	41.6
447	1,4-Benzenediol, 2-methyl-	95-71-6	41.4
448	Unknown		41.3
449	Unknown		41.3
450	1-Butanol, 2-methyl-	137-32-6	41.2
451	Decane, 2,4,6-trimethyl-	62108-27-4	41.0
452	Unknown		41.0
453	Unknown		40.9
454	Unknown		40.8
455	Unknown		40.7
456	Cyclopentanone, 2-methyl-	1120-72-5	40.5
457	Benzofuran	271-89-6	40.3
458	Unknown		40.3
459	Unknown		39.9
460	3-Penten-2-one, (E)-	3102-33-8	39.9
461	Unknown		39.3
462	Benzoic acid, 2,4-bis[(trimethylsilyl)oxy]-, trimethylsilyl ester	10586-16-0	39.0
463	Pyridine, 3-methyl-	108-99-6	38.6
464	Unknown 12713		38.6
465	Unknown 12717		38.5
466	Unknown		38.1
467	2-Buten-1-ol, 3-methyl-	556-82-1	38.0
468	Unknown		37.9
469	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	37.8
470	Unknown		37.6
471	2,3-Hexanedione	3848-24-6	37.6
472	Unknown		37.5
473	Nonane	111-84-2	37.2
474	Oxirane, butyl-	1436-34-6	37.1

475	Phenol, 4-pentyl-	14938-35-3	37.1
476	Unknown		37.0
477	Unknown		36.8
478	Ethanone, 1-(2-thienyl)-	88-15-3	36.7
479	Hexadecanoic acid, ethyl ester	628-97-7	36.5
480	Benzaldehyde, 3,4-dihydroxy-	139-85-5	36.5
481	Octadecanoic acid, ethyl ester	111-61-5	36.4
482	Unknown		36.3
483	2,4-Heptadienal, (E,E)-	4313-03-5	36.2
484	2-Acetyl-5-methylfuran	1193-79-9	36.2
485	Heptadecane	629-78-7	36.1
486	Unknown		36.1
487	2H-Pyran-2-one, tetrahydro-3,6-dimethyl-	3720-22-7	36.0
488	Unknown		35.8
489	Unknown		35.7
490	Octane, 4-methyl-	2216-34-4	35.6
491	Unknown		35.5
492	Cyclopropane, 1,1,2-trimethyl-3-(2-methyl-1-propenyl)-	54764-57-7	35.5
493	1H-Pyrrole-2-carboxaldehyde, 1-methyl-	1192-58-1	35.3
494	Unknown		35.3
495	Unknown		35.0
496	Unknown		35.0
497	1-Butanol, 3-methyl-, acetate	123-92-2	34.9
498	Unknown		34.8
499	Unknown		34.8
500	Unknown		34.8
501	2-Pentanone, 3-methyl-	565-61-7	34.7
502	Unknown		34.7
503	Unknown		34.6
504	Fampridine	504-24-5	34.6
505	Unknown		34.5
506	Unknown		34.4
507	Unknown		34.0
508	Unknown		33.9
509	Unknown		33.5
510	Unknown		33.4
511	2,5-Furandicarboxaldehyde	823-82-5	33.4
512	Unknown		33.4
513	2-Octanol	123-96-6	33.2
514	Unknown		33.2
515	Unknown		33.2
516	1-Heptene, 4-methyl-	13151-05-8	33.0
517	10-Methylundecan-4-olide	26346-34-9	32.9
518	Unknown		32.9
519	2,3-Pentanedione	600-14-6	32.5
520	Unknown		32.2
521	Unknown		32.2
522	Tetradecane	629-59-4	32.1
523	Unknown		32.1
524	Unknown		31.7
525	2-Cyclohexen-1-one, 3,4-dimethyl-	10463-42-0	31.7
526	1-Decen-3-yne	33622-26-3	31.6
527	Unknown		31.5

528	Cyclopropyl methyl carbinol	765-52-4	31.5
529	Unknown		31.4
530	Butanoic acid, methyl ester	623-42-7	31.4
531	Unknown		31.4
532	Butanal, 3-methyl-	590-86-3	31.3
533	Unknown		31.1
534	2-Cyclohexen-1-one, 3,5-dimethyl-	1123-09-7	30.9
535	Unknown		30.9
536	2-n-Butylacrolein	1070-66-2	30.7
537	1-Phenylpropene-3,3-diol diacetate	N/A	30.7
538	Unknown		30.5
539	Unknown		30.5
540	Ethylamine, N-methyl-N-hexadecyl-	66997-40-8	30.2
541	Unknown		30.1
542	Unknown		30.1
543	2(3H)-Furanone, dihydro-5-methyl-	108-29-2	29.9
544	Pentanal, 2-methyl-	123-15-9	29.8
545	Unknown		29.8
546	1,3-Nonadiene, (E)-	56700-77-7	29.6
547	Unknown		29.5
548	Unknown		29.4
549	Unknown		29.2
550	Unknown		29.1
551	2(5H)-Furanone, 3,5,5-trimethyl-	50598-50-0	29.0
552	Unknown		29.0
553	Unknown		28.8
554	2(3H)-Furanone, dihydro-5-pentyl-	104-61-0	28.7
555	Unknown		28.5
556	Unknown		28.2
557	Butanal, 2-methyl-	96-17-3	28.1
558	Pyrimidine, 2-methyl-	5053-43-0	27.5
559	Unknown		27.4
560	Isoamyl laurate	6309-51-9	27.2
561	2-Methyl-1-nonene-3-yne	70058-00-3	26.5
562	Unknown		26.4
563	Undecane, 2-methyl-	7045-71-8	24.9
564	Unknown		24.8
565	2,4-Dimethyl-1-heptene	19549-87-2	24.2
566	Unknown		23.7