Supporting Information

Development of an Enhanced Total Ion Current Chromatogram Algorithm to Improve Untargeted Peak Detection

Caitlin N. Cain, Sonia Schöneich, Robert E. Synovec*

Department of Chemistry, University of Washington, Box 351700, Seattle, WA, 98195-1700, USA

* Corresponding author:

Phone: +1-206-685-2328. Fax: +1-206-685-8665. E-mail: synovec@chem.washington.edu.

Table of Contents:

Table S1. List of analytes that made up the 90-component test mixture for the experimentally collected GC × GC-TOFMS chromatograms.[1]

Alkanes	**Alkynes**	**Esters**
Hexane	1-hexyne	Ethyl formate
Heptane	1-heptyne	Methyl decanoate
Octane	1-nonyne	Methyl caprylate
Nonane	5-decyne	Methyl salicylate
Decane	**Alcohols**	Ethyl salicylate
Undecane	1-propanol	Methyl laurate
Dodecane	2-butanol	Methyl caproate
Tridecane	1-pentanol	Diethyl phthalate
Tetradecane	2-pentanol	**Ketones**
Pentadecane	1-decanol	2-butanone
Hexadecane	1-tetradecanol	2-pentanone
Pristane	1-octadecanol	2-hexanone
Octadecane	Hexyl alcohol	3-hexanone
Eicosane	2-heptanol	2-heptanone
Halogenated Alkanes	1-octanol	3-heptanone
1,5-dichloropentane	1-nonanol	3-octanone
1-chlorohexane	1-eicosanol	2-decanone
1-bromohexane	Benzyl alcohol	2-undecanone
1-bromoheptane	2-ethyl-1-hexanol	2-dodecanone
1-bromooctane	**Aromatics**	**Aromatics**
1-chlorobutane	Benzene	1,2,4-trimethylbenzene
1,1,1-trichloroethane	Toluene	Anisole
1,2-dichloroethane	3-ethyltoluene	Dibutyl phthalate
Carbon tetrachloride	4-ethyltoluene	a-terpineol (90 %)
Cyclics	Mesitylene	
Methylcyclopentane	Ethylbenzene	
Cyclohexane	Butylbenzene	
Cyclooctane	Isobutylbenzene	
Butylcyclohexane	*t*-butyl benzene	
Bicyclohexyl	Propylbenzene	
2,2,4-trimethylpentane	1-ethylnapthelene	
Alkenes	Bromobenzene	
1-hexene	Cyclohexylbenzene	
Cyclohexene	Diphenylmethane	
Dodecane	p-xylene	
1-undecene	o-xylene	

Chromatographic Conditions

90-Component test mixture. Separations of a 90-component test mixture (Table S1) were collected on a GC × GC-TOFMS instrument configured with an Agilent 6890N GC (Agilent Technologies, Palo Alto, CA, USA) and LECO Pegasus III TOFMS (LECO Corporation, St. Joseph, MI, USA). A T-union was used to join the ^{1}D and ^{2}D columns to a pneumatic "pulse valve" (Model 009–0347–900, Parker Hannifin, Hollis, NH, USA), which acts as a total transfer modulator and operates using dynamic pressure gradient modulation (DPGM), a novel form of flow modulation. A Rtx-200 (19 m length, 180 µm d_c, 0.20 µm d_f) was the ^{1}D column and Rxi-1ms (2 m length, 180 µm d_c, 0.18 µm d_f) was the ^{2}D column (Restek, Bellefonte, PA, USA). The TOFMS collected data from 34 *m/z* to 338 *m/z*. The mass acquisition range applied was selected to ensure the capture of signal from all informative ions without interferences, while still maintaining a small data set computationally. The data collection rate was 200 Hz. The electron impact ionization energy was -70 eV and the detector voltage was 1562 V. The test mixture had an original concentration of 10 part-per-thousand (ppth), which was serially diluted in methanol to concentrations of 10 and 1 part-per-million (ppm). The chromatograms collected for the 10 and 1 ppm samples were utilized in the present study. The modulation period (P_M) was 1 s and the pulse width, defined by the length of time that flow is stopped from the pulse valve, was 200 ms. The oven temperature was initially held at 40 °C for 1 min and ramped to 250 °C (10 °C/min rate), where it was held for 1 min. The auxiliary pressure applied to the valve was held constant at 18.0 psig for 1 min and raised to 36.0 psig (0.857 psi/min rate), where it was held for 1 min. The 10 and 1 ppm samples were injected splitless in a 1 µL volume and the inlet flow rate was set to 1 mL/min. Additional experimental detail can be found in a previous report.[1]

Yeast cell metabolite extract. Furthermore, a GC × GC-TOFMS separation of yeast cells grown in respiring conditions was collected using an Agilent 6890N GC coupled to a LECO Pegasus III TOFMS with a 4D thermal modulator. Splitless injections in a 1 µL volume were injected and the flow rate was 1 mL/min. The ^{1}D column was a Rtx-5ms (20 m length, 250 µm d_c, 0.5 µm d_f) and the ^{2}D column was a Rtx-200 (2 m length, 180 µm d_c. and 0.2 µm d_f). The ^{1}D column was held at 60 °C for 0.25 min, increased to 280 °C at 8 °C/min, and was held at 280 °C for 10 min. The ^{2}D column followed the same temperature program except for its initial temperature of 70 °C. The P_M was 1.5 s and was kept 40 °C higher than the ^{1}D column. Data was collected at 100 Hz and the TOFMS collected data from 40 *m/z* to 600 *m/z*. More information describing the biological culture, extraction, and derivatization can be found in previous articles.[2,3]

Importing and Preprocessing of Experimentally Collected Chromatograms

Experimentally collected GC × GC-TOFMS chromatograms were imported into Matlab 2019b (Mathworks, Inc., Natick, MA, USA) using an in-house software, which converts the data from the ChromaTOF (LECO) format into Matlab variables.[3,4] Once chromatograms are unfolded into their vector form, baseline correction is performed on each *m/z* using a rolling ball, which subtracts low frequency noise from the data.[4,5] This technique operates as a "ball" with a fixed radius that rolls along the length of the chromatogram to extract the baseline.[4,5] The appropriate ball size must be chosen to preserve relevant chemical information and avoid overfitting the chromatogram. Optimization of the appropriate "ball radius" for baseline correction of a representative peak is shown in Figure S1. After the chromatograms are baseline corrected, their intensities were centered around zero. The chromatograms were re-folded, and the *m/z* dimension was summed together to create the standard total ion current chromatogram (TIC). A watershed-based algorithm[6,7] was performed for peak detection on the standard TIC using the Matlab Image Processing Toolbox. This approach for peak detection in the standard TIC was chosen because of its widespread use in commercially available software, such as Delta2D[4] and GC Image.[6,7]

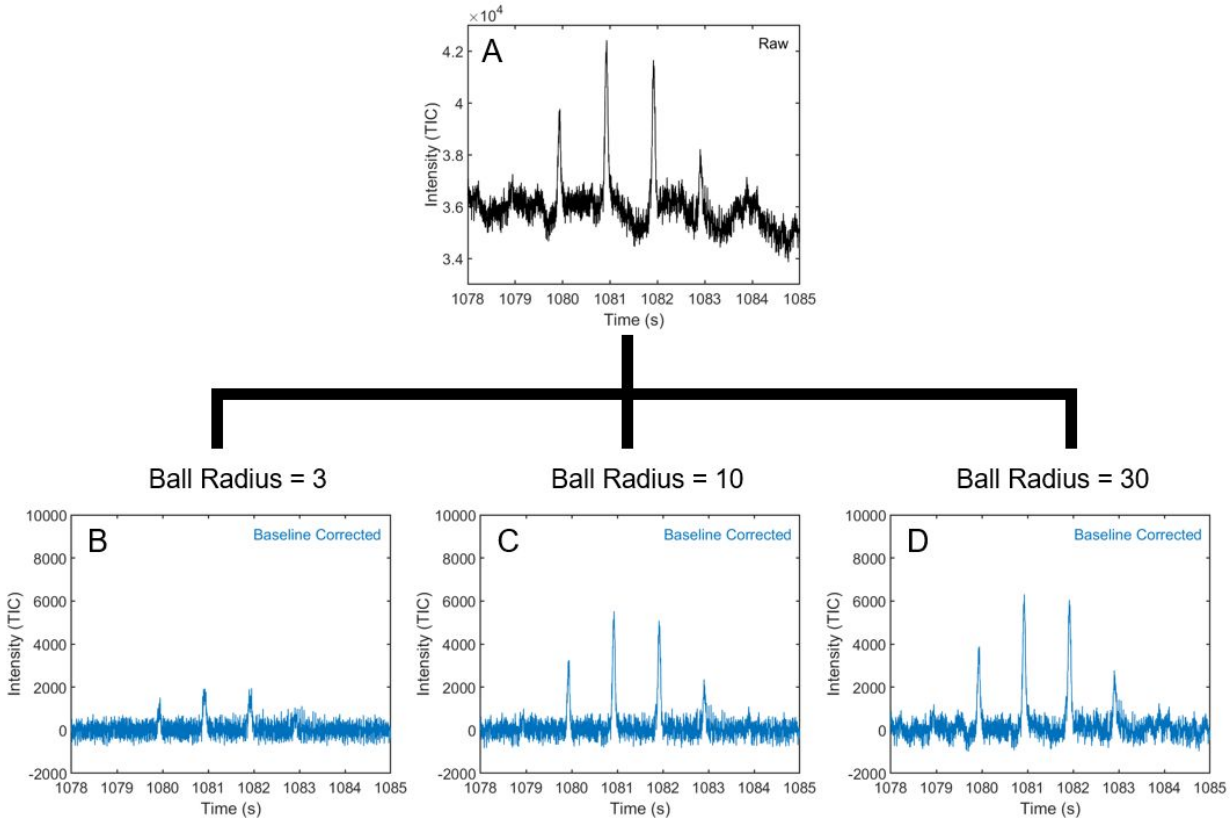

Figure S1. Schematic illustrating the optimization of the ball radius needed for baseline correction. This optimization process was used for baseline correction of the chromatograms of the 90-component test mixture and yeast cell extract metabolome. (A) The raw chromatogram for eicosane in the 10 ppm test mixture. (B) The chromatogram after applying a ball radius of 3 during baseline correction. This ball radius was determined to be too small since it overfitted the baseline. (C) The chromatogram after applying a ball radius of 10 during baseline correction. The ball radius was determined to be appropriate for baseline correction. (D) The chromatogram after applying a ball radius of 30 during baseline correction. This ball radius was determined to be too large since it did not fully subtract out the low frequency noise in the baseline.

Table S2. Chromatographic parameters used in GC × GC-TOFMS simulations.

Parameter	**Conditions studied**
^{1}D separation time, $^1t_{sep}$ (s)	160, 80, 53, 40, 32, 27, 23, 20, 18, 16
Modulation period, P_M (s)	1
Number of components, m	40
Saturation factor, α_{2D}	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1
Average peak area (TIC)	200
^{1}D peak width-at-base, ^1w_b (s)	4
^{2}D peak width-at-base, ^2w_b (ms)	100
Data collection rate (Hz)	100
S/N relative to the average peak area, S/N_{rel}	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, 5, 10, 15, 20, 50, 100

Table S3. List of the analytes selected for the simulation study. Abbreviations: MEOX – methoximation derivatization; TMS – trimethylsilylation derivatization.

Allantoin, 3 TMS	L-Cysteine, 3 TMS	D-(+)-Glucosamine, 6 TMS	L-Isoleucine, TMS	Phenylalanine, 2 TMS
2-Aminoadipic acid, 3 TMS	Cytosine, N,O-2 TMS	D-Glucose, 5 TMS	Isomaltose, TMS	Phosphoric acid, TMS
4-Aminobutanoic acid, 3 TMS	1,3-Diaminopropane, 4 TMS	Glucuronic acid, TMS	Lactic Acid, 2 TMS	L-Proline, 2 TMS
Arabinose, MEOX 4 TMS	Dopamine, 3 TMS	L-Glutamine, 3 TMS	Leucine, 2 TMS	Putrescine, 4 TMS
Aspartic acid, 3 TMS	Erythritol, 4 TMS	Glyceraldehyde, MEOX 2 TMS	Lysine, 3 TMS	Pyruvic acid, MEOX TMS
Benzoic acid, TMS	β-D(-)-Erythrose, TMS	Glyceric acid, 3 TMS	Malic acid, 3 TMS	α-Ribose, TMS
β-Alanine, 2 TMS	Ethanolamine, 3 TMS	Glycine, 3 TMS	Methylcysteine, TMS	Threonine, 3 TMS
Butyric Acid, TMS	Ethylene glycol, 2 TMS	Glycolic acid, 2 TMS	Methylmalonic acid, 2 TMS	Thymine, 2 TMS
Cadaverine, 3 TMS	Ferulic acid, 2 TMS	Glyoxylic acid, MEOX TMS	Norepinephrine, 4 TMS	α-Tocopherol, TMS
Caffeic acid, 3 TMS	D-Fructose, 5 TMS	Hexanoic acid, TMS	L-Norleucine, 2 TMS	Tyramine, 2 TMS
Campesterol, TMS	L-Fucose, 4 TMS	Histidine, N,N,O-3 TMS	Normetanephrine, 3 TMS	L-Tyrosine, 2 TMS
Cholesterol, TMS	Fumaric acid, 2 TMS	4-Hydroxybenzoic acid, 2 TMS	L-Norvaline, 2 TMS	Urea, 2 TMS
Citric acid, 3 TMS	Galactitol, 6 TMS	L-Hydroxyproline, N,O,O-TMS	L-Ornithine, 4 TMS	L-Valine, 2 TMS
Citrulline, TMS	Galactose, TMS	Hydroxypyruvic acid, MEOX 2 TMS	2-Oxobutyric acid, MEOX TMS	Xylitol, 5 TMS
Cystathionine, 4 TMS	D-Gluconic acid, 6 TMS	3-Indoleacetic acid, 2 TMS	L-5-Oxoproline, 2 TMS	Xylose, 4 TMS

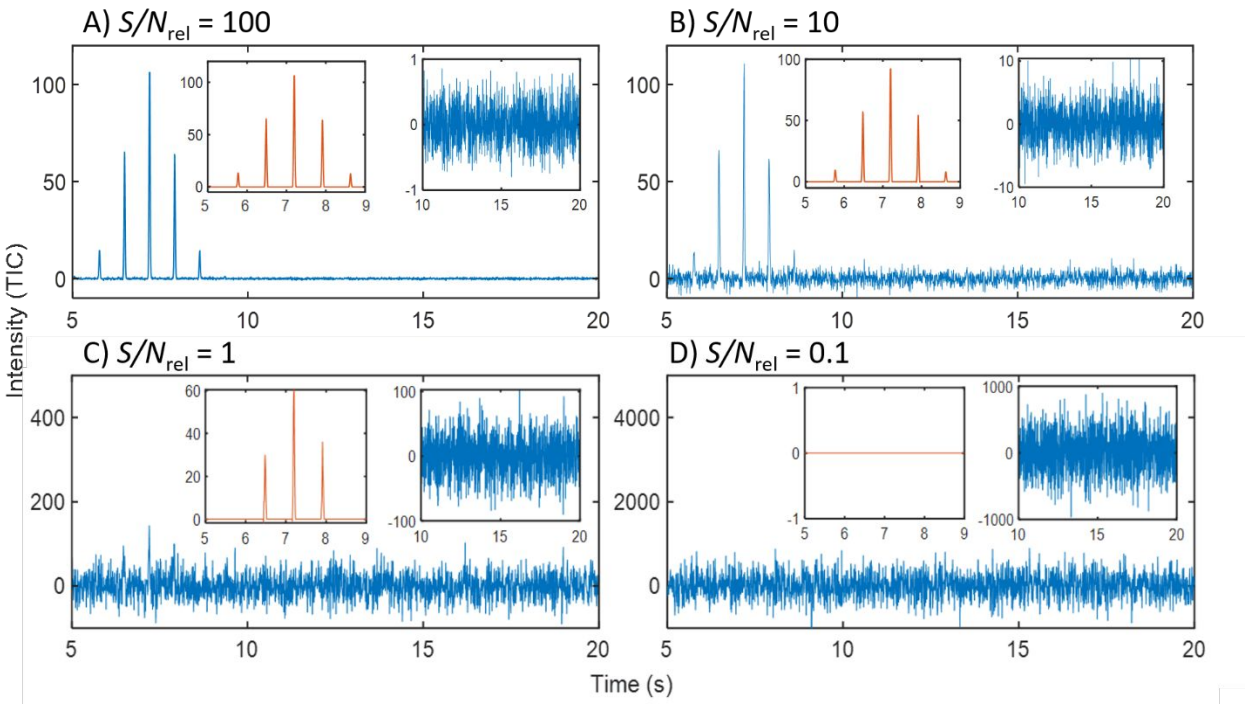

Figure S2. A simulated unfolded peak with an area of 200 at four different S/N_{rel} values: (A) 100, (B) 10, (C) 1, and (D) 0.1. The left inset in each panel shows the enhanced TIC version of the peak (red) and the right inset zooms in on the baseline noise.

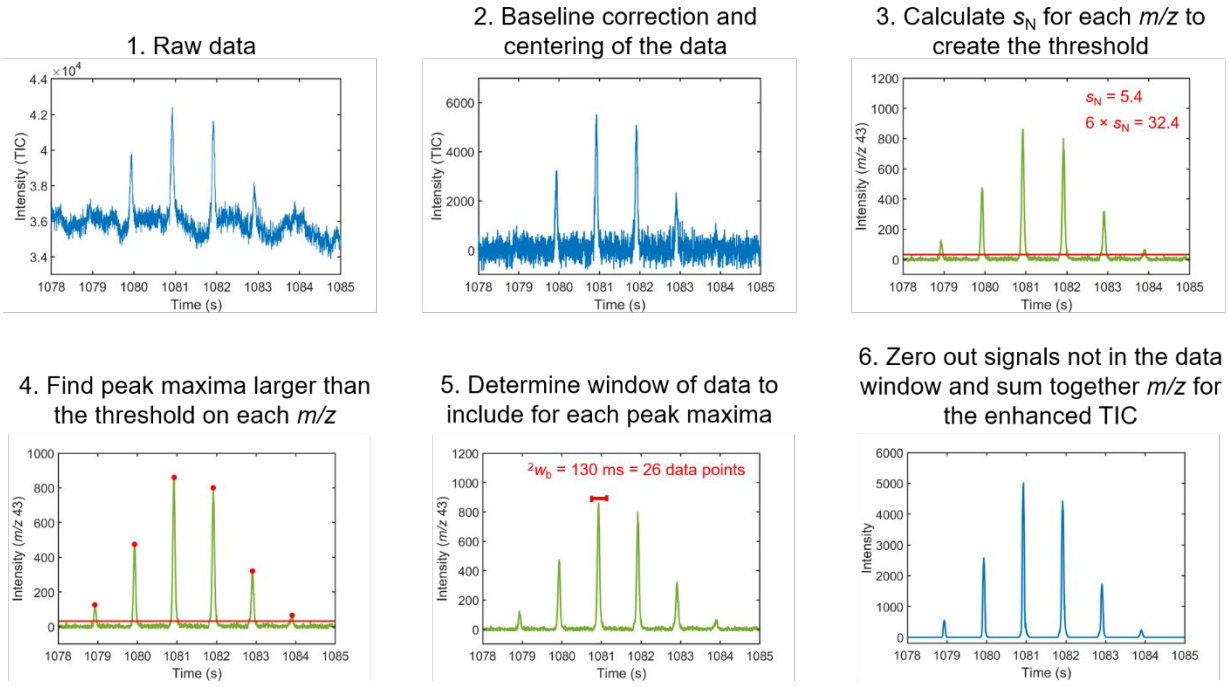

Figure S3. Schematic illustrating each step of the enhanced TIC algorithm (see Experimental Section) on eicosane in the 10 ppm 90-component test mixture. See Figures S4 and S5 for more details regarding selection of the appropriate signal threshold in Step 3.

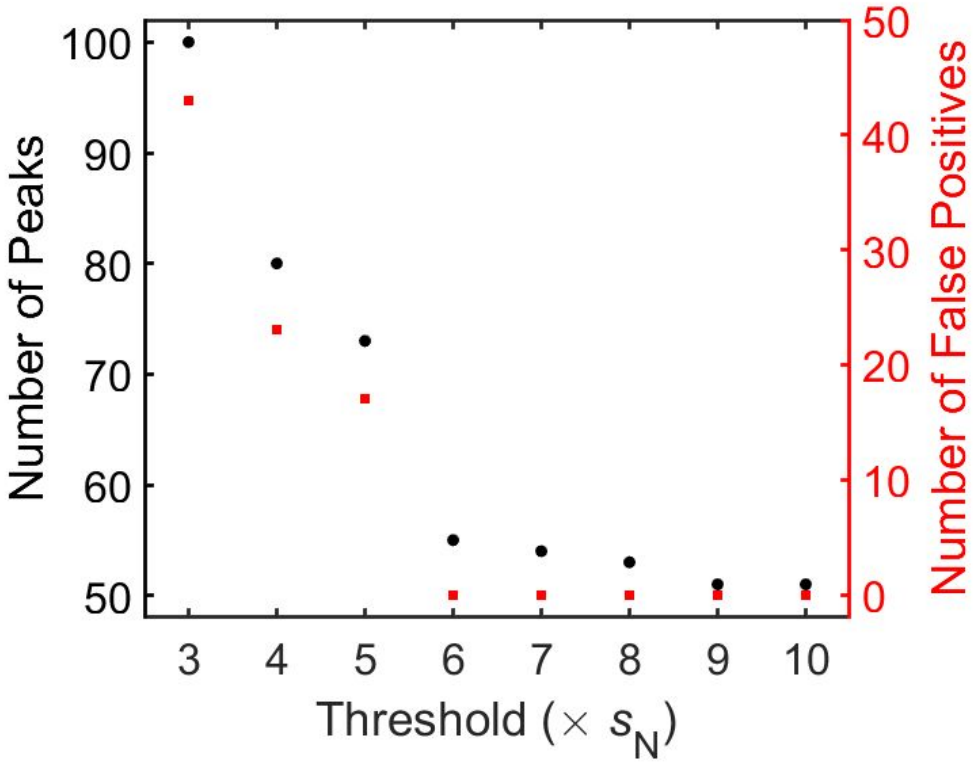

Figure S4. The number of peaks (black circles) and number of false positives (red squares) as a function of the signal threshold applied during the enhanced TIC method on the 10 ppm test mixture.

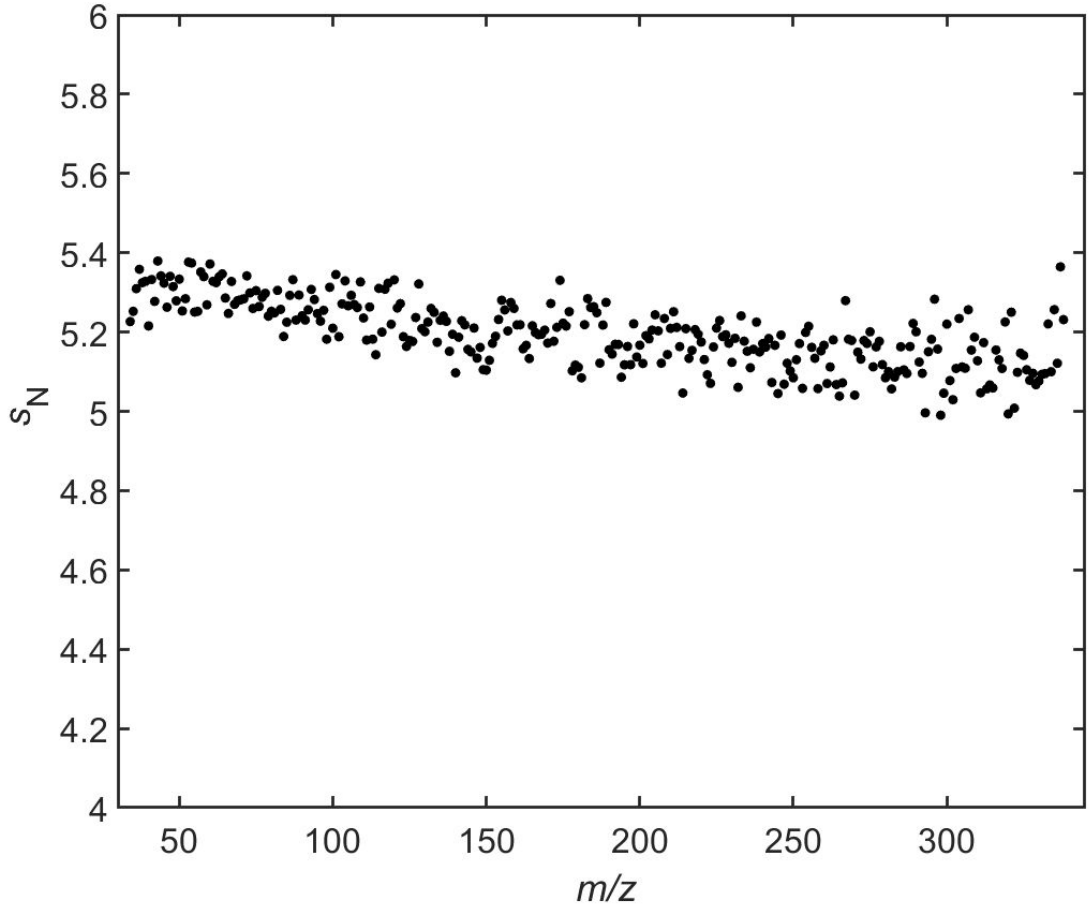

Figure S5. The standard deviation of the baseline noise (s_N) on each mass channel, *m/z*, in the 10 ppm 90-component test mixture.

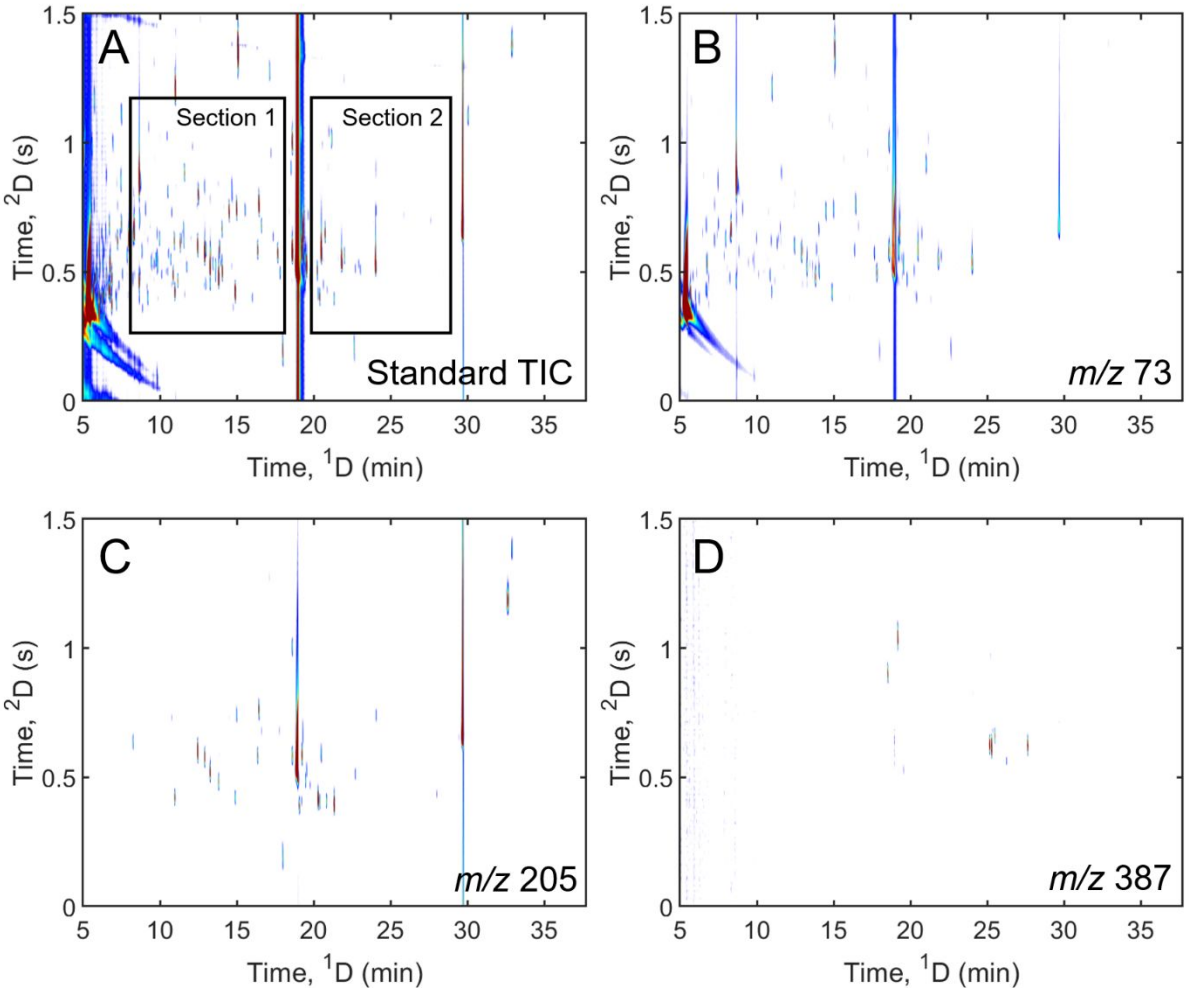

Figure S6. GC × GC-TOFMS separation of a metabolite extract collected from respiring yeast cells, metabolizing ethanol. (A) The TIC produced by summing the mass spectral dimension after baseline correction. The two boxes labeled as Section 1 and Section 2 correspond to time windows highlighted in Figure 4. (B) The extracted ion current chromatogram (EIC) of the separation in (A) at *m/z* 73. (C) The EIC of the separation in (A) at *m/z* 205, which is selective towards carbohydrates. (D) The EIC of the separation in (A) at *m/z* 387, which is selective towards sugar phosphates.

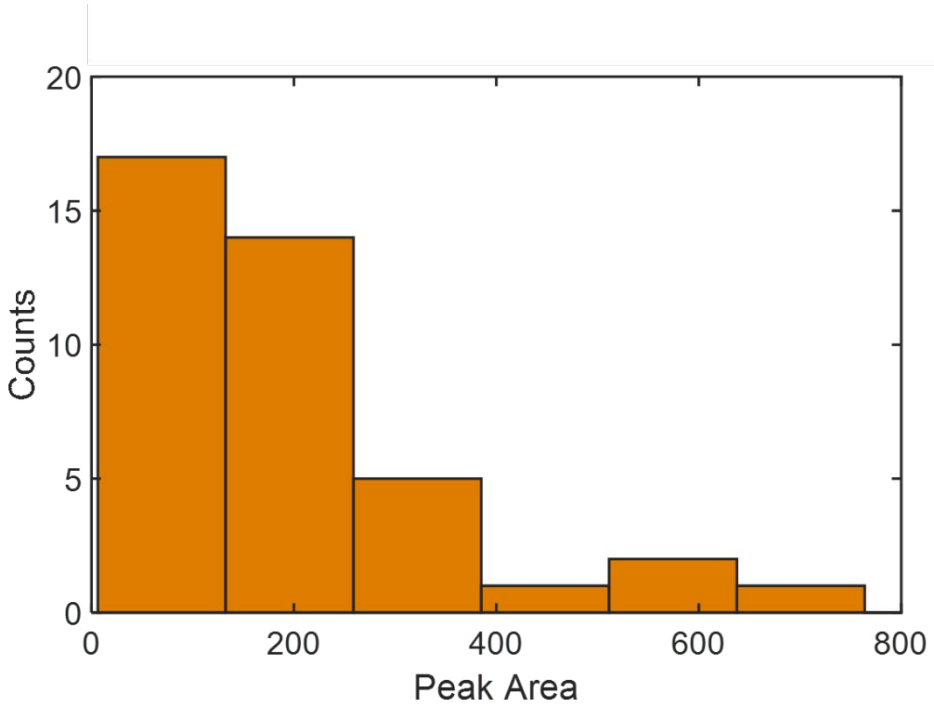

Figure S7. Exponential distribution of peak areas used for the representative simulations in Figure 6.

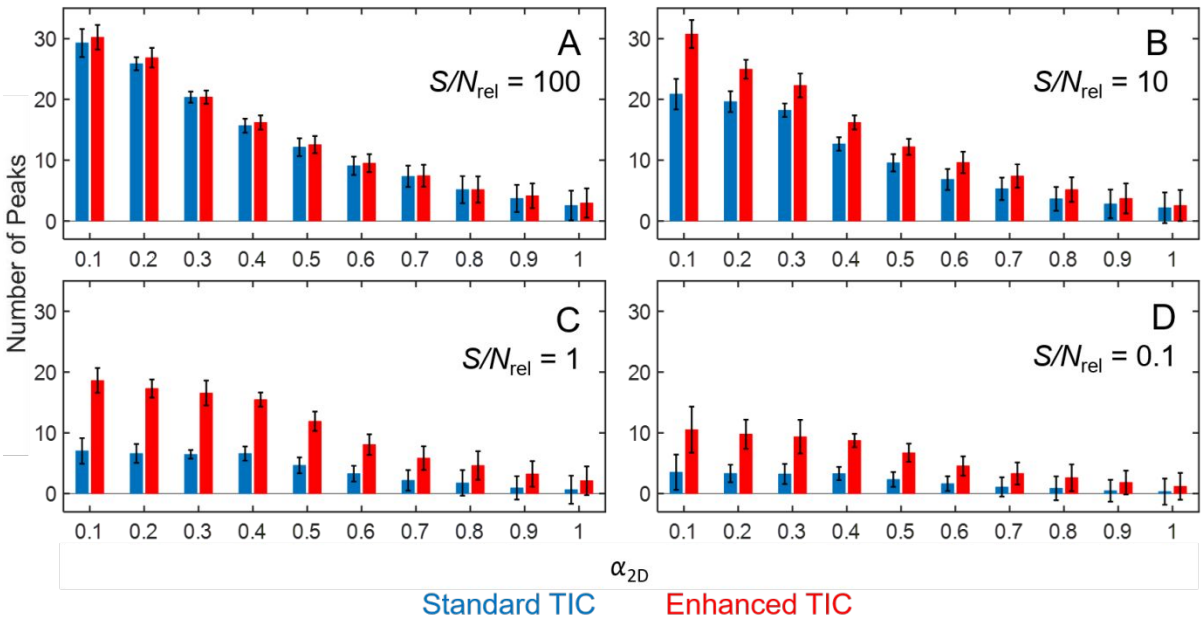

Figure S8. The number of peaks detected for both the standard TIC (blue) and enhanced TIC (red) as a function of the saturation factor (α_{2D}) simulated at four different S/N_{rel} values: (A) 100, (B) 10, (C) 1, (D) 0.1. Results are shown as the average and standard deviations of 100 simulations for each S/N_{rel} value.

Table S4. Comparison of the lack of fit (%) between SOT (eq 7) and the simulated chromatographic results for the standard and enhanced TICs at different S/N_{rel} in Figure 7.

S/N_{rel}	Standard TIC	Enhanced TIC
100	5.0	2.5
10	25.2	5.2
1	68.0	20.0
0.1	83.9	53.4

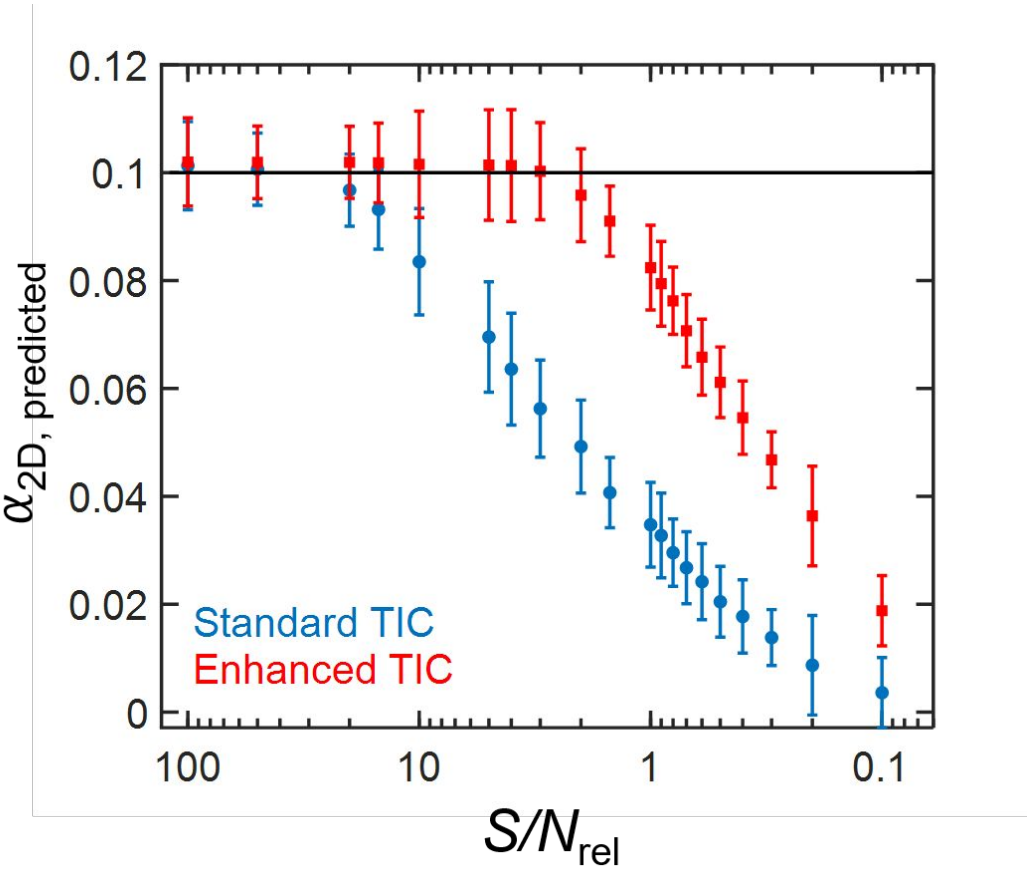

Figure S9. The effect of S/N_{rel} on the $\alpha_{2D,predicted}$ using the statistical overlap theory for both the standard TIC (blue circles) and enhanced (red squares) TIC. The true α_{2D} for all the simulations was 0.1 (40 analytes in a peak capacity space ($n_{c,2D}$) of 400). Results are shown as the average and standard deviations of 100 simulations.

References

(1) Schöneich, S.; Gough, D. V.; Trinklein, T. J.; Synovec, R. E. Dynamic Pressure Gradient Modulation for Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Detection. *J. Chromatogr. A* **2020**, *1620*, 460982. https://doi.org/10.1016/j.chroma.2020.460982.

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

(3) Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Pierce, K. M.; Young, E. T.; Synovec, R. E. Comprehensive Analysis of Yeast Metabolite GC×GC-TOFMS Data: Combining Discovery-Mode and Deconvolution Chemometric Software. *Analyst* **2007**, *132* (8), 756–767. https://doi.org/10.1039/b700061h.

(4) Kehimkar, B.; Hoggard, J. C.; Marney, L. C.; Billingsley, M. C.; Fraga, C. G.; Bruno, T. J.; Synovec, R. E. Correlation of Rocket Propulsion Fuel Properties with Chemical Composition Using Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry Followed by Partial Least Squares Regression Analysis. *J. Chromatogr. A* **2014**, *1327*, 132–140. https://doi.org/10.1016/j.chroma.2013.12.060.

(5) Schmarr, H. G.; Bernhardt, J. Profiling Analysis of Volatile Compounds from Fruits Using Comprehensive Two-Dimensional Gas Chromatography and Image Processing Techniques. *J. Chromatogr. A* **2010**, *1217* (4), 565–574. https://doi.org/10.1016/j.chroma.2009.11.063.

(6) Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A. Information Technologies for Comprehensive Two-Dimensional Gas Chromatography. *Chemom. Intell. Lab. Syst.* **2004**, *71* (2), 107–120. https://doi.org/10.1016/j.chemolab.2003.12.009.

(7) Reichenbach, S. E.; Tian, X.; Tao, Q.; Stoll, D. R.; Carr, P. W. Comprehensive Feature Analysis for Sample Classification with Comprehensive Two-Dimensional LC. *J. Sep. Sci.* **2010**, *33* (10), 1365–1374. https://doi.org/10.1002/jssc.200900859.