Supporting Information Section for:

A Systematic Method for the Targeted Discovery of Chemical Attribution Signatures: Application to Isopropyl Bicyclophosphate Production

Carolyn L. Mazzitelli; Michael A. Re; Melissa A. Reaves; Carlos A. Acevedo; Stephen D. Straight; and Joseph E. Chipuk*

Signature Science, LLC., 8329 North MoPac Expressway, Austin, TX, 78759

ABSTRACT

The Supporting Information contains details for the synthesis of **Triol** and IPBCP via Routes 1, 2 3, 4, and 5. Table S-1 showing the experimental design as well as experimental details for sample preparation UPLC-MS method conditions, system suitability criteria, data conversion, and data processing with a summary shown in Scheme S-1 are also included. Additional APCI - mass spectra and LC-MS ion chromatograms detailing the identification of potential signatures are included in Figures S-1 and S-2.

SUPPORTING INFORMATION

Synthesis of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol (Triol). A modification of the method described by Derfer *et al* was used.[20] In summary, 11.6 g (0.29 mol) of sodium hydroxide was added to a 500 mL reaction flask followed by the addition of 40 mL of water. After the sodium hydroxide completely dissolved, 140 mL (1.85 mol) of 36.5% formaldehyde was added to the flask. After stirring for approximately 5 min, isovaleraldehyde (20.0 mL, 0.19 mol) was added slowly over approximately 10 min. The reaction was heated to reflux and stirred for approximately 15 h during which time the solution became slightly cloudy/white. The reaction was cooled to room temperature, 200 mL of ethyl acetate was added, and the solution was stirred rapidly for approximately 1 hr. The ethyl acetate layer was separated. Sodium chloride was added to the aqueous layer until saturation. The aqueous layer was then extracted with three 100 mL aliquots of ethyl acetate. The organic fractions were combined, dried over sodium sulfate, and concentrated via rotary evaporation to yield a clear, viscous oil. The oil was distilled under reduced pressure to yield a thick oil (10.5 g) that solidified upon cooling. Finally, the material was recrystallized from hexane:ethyl acetate to yield a white crystalline solid (7.5 g, 27%). APCI-MS: *m/z* 149, $[M + H]^+$; HRMS: *m/z* calculated for $C_7H_{17}O_3$: 149.1178, found at 149.1173; ^{1}H NMR: δ 0.88 [6 H, d], 1.75 [1H, m], 3.10 [3H, br s],3.81 [6H, s].

Chemical Attribution Signature Study Synthesis of IPBCP using Phosphorous Oxychloride (Route 1). For each reaction, 25 mg of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol was added to an 8 mL reaction vial, followed by 0.5 mL of the specified reaction solvent. To the stirring mixture, pyridine (44 µL, 3 molar equivalents) was added, followed by phosphorous oxychloride (15 µL, 1 molar equivalent) and the vial was sealed with a septa cap. The reaction was heated at 50 °C in an aluminum heating block for 24 hrs with

constant stirring and then cooled to ambient temperature. The solvent level was checked and, if necessary, additional solvent was added to bring the final reaction volume to 0.5 mL. For reactions performed without solvent, 0.5 mL of 50:50 acetonitrile:water was added to the reaction vial. All reactions were diluted with an additional 100 µL aliquot of 50:50 acetonitrile:water.

Chemical Attribution Signature Study Synthesis of IPBCP using Phosphoric Acid (Route 2). For each reaction, 25 mg of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol was added to an 8 mL reaction vial, followed by 0.5 mL of the specified reaction solvent. To the stirring mixture, phosphoric acid (35 µL, 2 molar equivalents) was added to the vial and the vial was sealed with a septa cap. The reaction was heated at 100 °C in an aluminum block for 24 hrs with constant stirring and then cooled to ambient temperature. The solvent level was checked and, if necessary, additional solvent was added to bring the final reaction volume to 0.5 mL. For reactions performed without solvent, 0.5 mL of 50:50 acetonitrile:water was added to the reaction vial. All reactions were diluted with an additional 100 µL aliquot of 50:50 acetonitrile:water.

Chemical Attribution Signature Study Synthesis of IPBCP through the bicyclophosphite intermediate using Phosphorous Pentachloride (Route 3). For each reaction, 25 mg of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol was added to an 8 mL reaction vial, followed by 0.5 mL of the specified reaction solvent. To the stirring mixture, pyridine (73 µL, 5 molar equivalents) was added, followed by phosphorous pentachloride (38 µL, 1 molar equivalent) and the vial was sealed with a septa cap. The reaction was heated at 50 °C in an aluminum block for 24 hrs with constant stirring and then cooled to ambient temperature. The solvent level was checked and, if necessary, additional solvent was added for a final volume of 0.5 mL. For

reactions performed without solvent, 0.5 mL of 50:50 acetonitrile:water was added to the reaction vial. All reactions were diluted with an additional 100 µL aliquot of 50:50 acetonitrile:water.

Chemical Attribution Signature Study Synthesis of IPBCP through the bicyclophosphite intermediate using Triethyl Phosphite and Hydrogen Peroxide (Route 4). For each reaction, 25 mg of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol was added to an 8 mL reaction vial, followed by 0.5 mL of the specified reaction solvent. To the stirring mixture, pyridine (30 µL, a catalytic amount) was added, followed by triethyl phosphite (65 µL, 1 molar equivalent) and the vial was sealed with a septa cap. The reaction was heated at 50 °C in an aluminum block for 24 hrs with constant stirring and then cooled to ambient temperature. The solvent level was checked and, if necessary, additional solvent was added for a final volume of 0.5 mL. After bringing the volume to 0.5 mL (if needed), 100 µL of hydrogen peroxide was added to the vial and the mixture was stirred for approximately 10 minutes. If the reaction was performed without solvent, 0.5 mL of 50:50 acetonitrile:water was added to the reaction vial.

Chemical Attribution Signature Study Synthesis of IPBCP through the bicyclophosphite intermediate using Phosphorous Trichloride and Hydrogen Peroxide (Route 5). For each reaction, 25 mg of 2-(hydroxymethyl)-2-(propan-2-yl)propane-1,3-diol was added to an 8 mL reaction vial, followed by 0.5 mL of the specified reaction solvent. To the stirring mixture, pyridine (44 µL, 3 molar equivalents) was added, followed by phosphorous trichloride (16 µL, 1 molar equivalent) and the vial was sealed with a septa cap. The reaction heated at 50 °C in an aluminum block for 24 hrs with constant stirring and then cooled to ambient temperature. The solvent level was checked and, if necessary, additional solvent was added for a final volume of 0.5 mL. After bringing the volume to 0.5 mL (if needed), 100 µL of hydrogen peroxide was

added to the vial and the mixture was stirred for approximately 10 minutes. If the reaction was performed without solvent, 0.5 mL of 50:50 acetonitrile:water was added to the reaction vial.

Table S-1. Experimental design showing the number of samples synthesized per specified reaction condition.

Phosphorous Reagent	**Phosphorous Reagent Manufacturer**	**Solvent**				**Total Samples**
		Acetone	Ethyl Acetate	None	Acetonitrile	
PCl_3	Strem	3	3	3	3	12
	Aldrich	3	3	3	3	12
	Alfa Aesar	3	3	3	3	12
H_3PO_4[a]	Aldrich	3	3	3	3	12
	Acros	3	3	3	3	12
	JT Baker	3	3	3	3	12
$P(OEt)_3$	Strem	3	3	3	3	12
	Alfa Aesar	3	3	3	3	12
	TCI	3	3	3	3	12
$POCl_3$	Strem	3	3	3	3	12
	Alfa Aesar	3	3	3	3	12
	Acros	3	3	3	3	12
PCl_5	Strem	3	3	3	3	12
	Aldrich	3	3	3	3	12
	Alfa Aesar	3	3	3	3	12
Total Samples		45	45	45	45	180

[a]For H_3PO_4 reactions, temperature is constant at 100°C.

Sample Preparation. Following reaction completion, solutions were brought to a total volume of 0.6 ml in 8 ml scintillation vials. The solutions were then vortexed for approximately 10 seconds. A 100 µL aliquot of the reaction solution was removed and subsequently diluted to 1 mL total volume with a diluent comprised of 50:50 acetonitrile:water. All samples were spiked with 2.5 ppm of rhodamine B, which served as the internal standard for the LC-MS analysis.

Method blanks containing the internal standard spiked into the diluent (50:50 acetonitrile:water) and solvent blanks that contained the diluent only were prepared on each day of analysis. To insure that the impurity profile of the solvents used to prepare the samples matched those of the solvent blanks and method blanks, the aliquots for the method blanks and solvent blanks were taken from the same extraction solvents used to prepare the samples. In addition, reaction blanks containing the reaction solvent only were also prepared and subjected to the same reaction conditions and sampling procedure as the actual samples.

UPLC-MS Method Conditions. Mobile phase A was comprised of 0.1% formic acid in water (*v/v*) and Mobile Phase B contained 0.1% formic acid in acetonitrile. The following gradient was used: at a flow rate of 0.45 mL/min, the initial conditions of 98% A/2% B were held for 1.0 min, followed by a linear gradient to change the composition to 25.5% A/74.5% B over 13 minutes. The eluent make-up was ramped to 5% A/95% B over 0.25 minutes and held for an additional 2.0 min. The gradient was restored to initial conditions in 0.25 min, followed by a 2.0 min re-equilibration time. A reversed-phase Hypersil Gold UPLC column, 50 × 2.1 mm, 1.7 µm (Thermo Scientific, San Jose, CA) heated to 40 ºC was used. Detection was by UV using an Accela photodiode array (PDA) detector (Thermo Scientific, San Jose, CA) and an in-line Thermo Fisher Scientific LTQ linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA).

Ionization was achieved using positive ion mode APCI with a source current of 5 µA and a vaporizer temperature of 450 ºC. The sheath and auxiliary gas flow rates were 50 and 5 arbitrary units, respectively. The heated inlet capillary temperature was 275 ºC. The linear ion trap was operated in the full scan mode, and spectra were collected for the range of *m/z* 50 to *m/z* 1000. For targeted collision induced dissociation (CID) and MS^n, as well as data dependent MS^n experiments, an isolation width of 3.0 and normalized collision energy of 35% was used for ions not containing the chlorine isotope pattern. For MS^n experiments performed on compounds suspected of containing chlorine ions, an isolation width of 6.0 was used. For high resolution mass spectrometry experiments, spectra were collected for the range of *m/z* 50 to *m/z* 1000 with a resolution of 50,000 and an AGC target of 1×10^6.

System Suitability. The following system suitability criteria were specified prior to the start of the study and met for each analysis: a) any peaks present in the method blank injections must not be greater than 1% of the peak area of the internal standard. Peaks greater than 1% were flagged as possible contaminants, b) the percent relative standard deviation (%RSD) of the internal standard peak area in the first five injections of the method blank must not be larger than 20%, c) the retention time of the internal standard in the sample injections and the bracketing method blank injections must be within ± 1.0 min of the mean retention time of the internal standard in the first five method blank injections, d) the percent recovery of the internal standard in the sample injections and the bracketing method blank injections must be within 50 to 150% relative to the mean of the first five method blank injections, e) carryover of the IPBCP peak must be evaluated in the chromatograms of the method blank samples bracketing the sample injections. If the IPBCP peak was present, its peak area must not be greater than 1% of the peak area of the internal standard.

Data Conversion. The LC-MS data files, acquired as *.raw files using the Thermo Fisher XCalibur instrument control software, were loaded into a database prior to further manipulation. The Xcalibur Development Kit (XDK) enabled the extraction of data from the XCalibur raw files using programmable COM objects. By incorporating the COM objects in a custom Visual Basic script (Microsoft Corporation, Redmond, WA), the raw data (RT, *m/z*, intensity) and metadata, were systematically extracted without further manual preprocessing of the sample. Each sample was then extracted and imported into a Microsoft Access database (Microsoft Corporation, Redmond, WA) to better organize the data and to perform initial data reduction prior to data manipulation and analysis. To reduce the size of the data files, *m/z* values present at less than 0.15% of the maximum observed intensity at a specific retention time were removed from each file. After this filtering process was completed, the samples were loaded into an Oracle 11g database (Oracle Corporation, Redwood Shores, CA) via an ODBC connection for further data processing and analysis.

Data Analysis. SAS 9.2 (SAS, Cary, NC), TIBCO Spotfire S+ 8.2 (TIBCO, Palo Alto, CA) and R-project (R Foundation, Vienna, Austria) were used for data analysis, additional data reduction, filtering, and generation of interactive scatter plots. All of the essential processing was performed using the SAS system after the Oracle database was imported using specific functions and algorithms that were built into the code to handle each of the steps. TIBCO and R-project software were used to display the potential signatures through a series of interactive 2D and 3D plots.

A summary of the data analysis process is shown in Scheme S-1 in the Supporting Information section. The first five method blank injections from an injection sequence were used to identify noise, background, and artifact ions not associated with components of the sample. These blanks

were processed as follows: A raw data matrix consisting of a plot of all data points with *m/z* plotted on the y-axis, retention time (RT) plotted on the x-axis, and intensity plotted on the z-axis was generated from all data files. The retention time of the internal standard was then used to calculate the relative retention time (RRT) for all ions in a sample to normalize run-to-run variability in retention time. The RRT was calculated by dividing the retention time of each component of the sample by the retention time of the internal standard as shown below:

$$RRT_{component} = RT_{component} / RT_{rhodamine\ B\ internal\ standard}$$

The RRT was used in conjunction with the *m/z* values to construct a matrix of sets, defined as a specific *m/z* value with its relative retention time and an absolute intensity. This matrix was generated for each data file by binning the data points by units of 1 *m/z* on the y-axis and 0.01 for RRT on the x-axis. If multiple ion intensities of the same *m/z* were observed for the same binned RRT, those intensities were summed to obtain a unique intensity value for that set. The noise and background threshold for each analysis sequence was then calculated by determining the average intensities for sets observed in the first five method blanks run in a sequence and multiplying this value by nine.

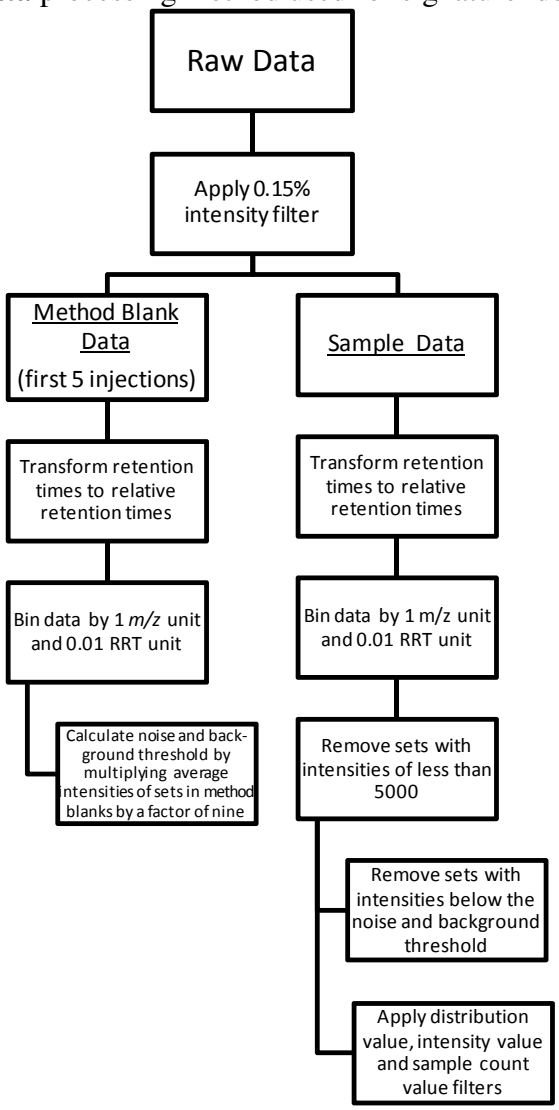
As shown in Scheme S-1, sample files were processed by first generating data matrices using the same procedure described above for the method blanks. Sets with intensity values of less than 5000 and sets containing intensities below the noise and background threshold were then filtered out of the data matrices to ensure the resulting data matrices contained only the sets unique to the samples. To identify the sets that were potential chemical attribution signatures, groups of samples, referenced here as classes, were compared. For example, when evaluating the route-dependent signatures, five classes were compared, each comprised of only the relevant samples

from each of the five synthetic routes. Only sets that occurred in at least 2 samples within a class and that were unique to a particular class within ±0.02 RRT were considered potential signatures.

Data filters were established to prioritize potential signatures using the following values: total intensity value, percent intensity value, and sample count value. The total intensity value (the sum of the intensities of all data points for a set) is an indicator of the intensity of a particular set. The total intensity value was used to rank the sets, such that sets with a higher total intensity value were given higher priority during final signature evaluation than those with a lower total intensity value.

The percent intensity value was used to provide an indication of how intense a set was in a specific class, relative to the total observed intensity of the set in all samples. The intensity for a set within a class was divided by the sum of intensities for that set across all samples included in the experimental design. The sample count value was also determined to indicate the number of samples for which a unique set was observed within a class. For example, if a unique set was seen in five samples and all five samples came from the same class, then it is very likely that the set is a signature of that class. An additional requirement for signature identification was the presence of the compound in at least two out of the three reaction replicates for each set of samples. Conversely, for the signature to be considered absent in a class, it must not be present in more than one of the three reaction replicates. Upon identification of a potential attribution signature, its presence in all samples included in the experimental design was also assessed to ensure it was unique to a specific synthetic route, source, condition, or combination thereof.

Scheme S-1. Summary of the **data** processing method used for signature identification



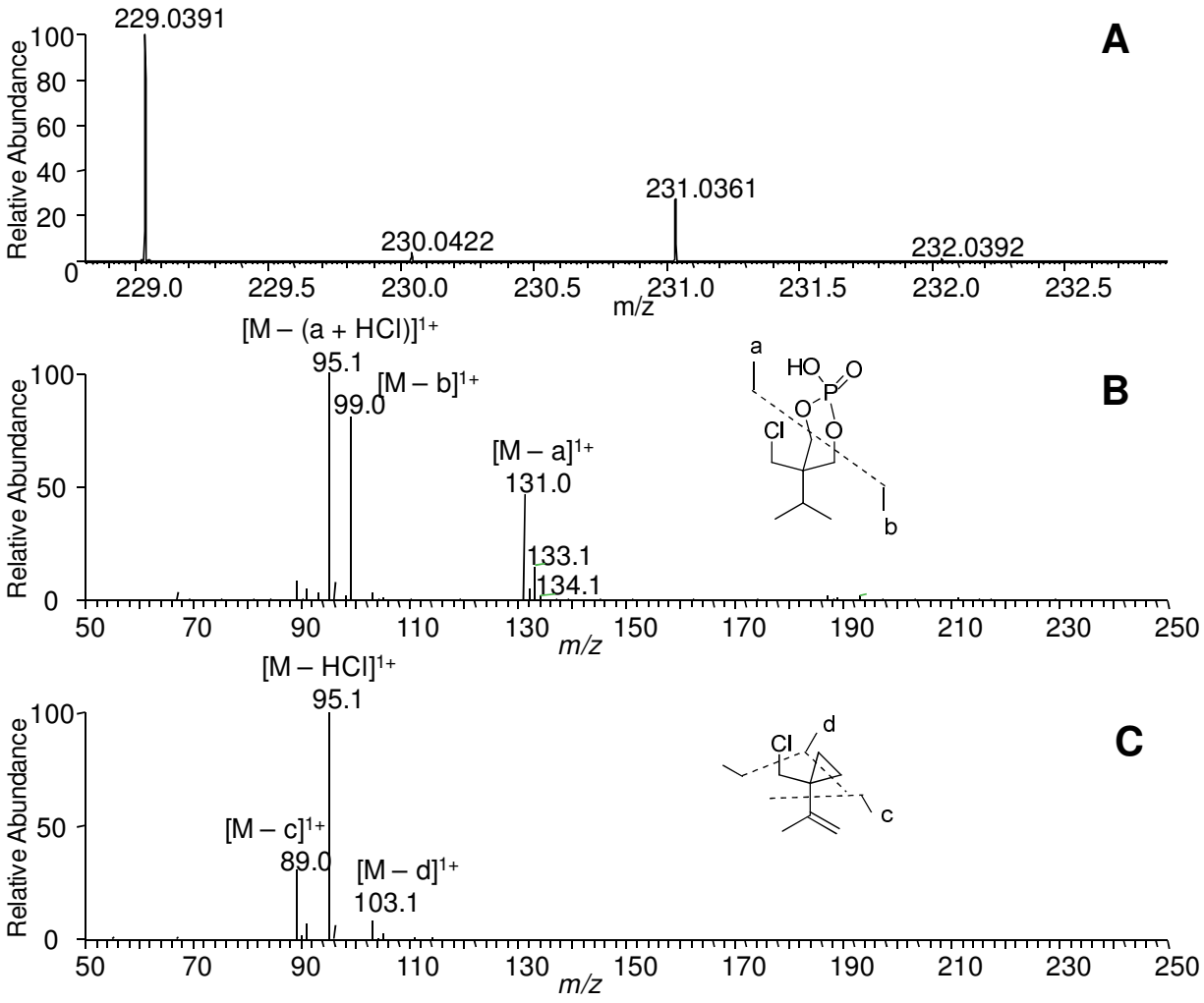


Figure S-1. APCI-mass spectra for *m/z* 229 at RRT 1.16 in an attribution study sample synthesized by Route 3 (RDEN0131): (A) High resolution mass spectrum showing the exact mass and isotopic distribution of *m/z* 229, (B) MS/MS spectrum of *m/z* 229, and C) MS^3 spectrum for *m/z* 229 → 131 $[M - a]^{1+}$.

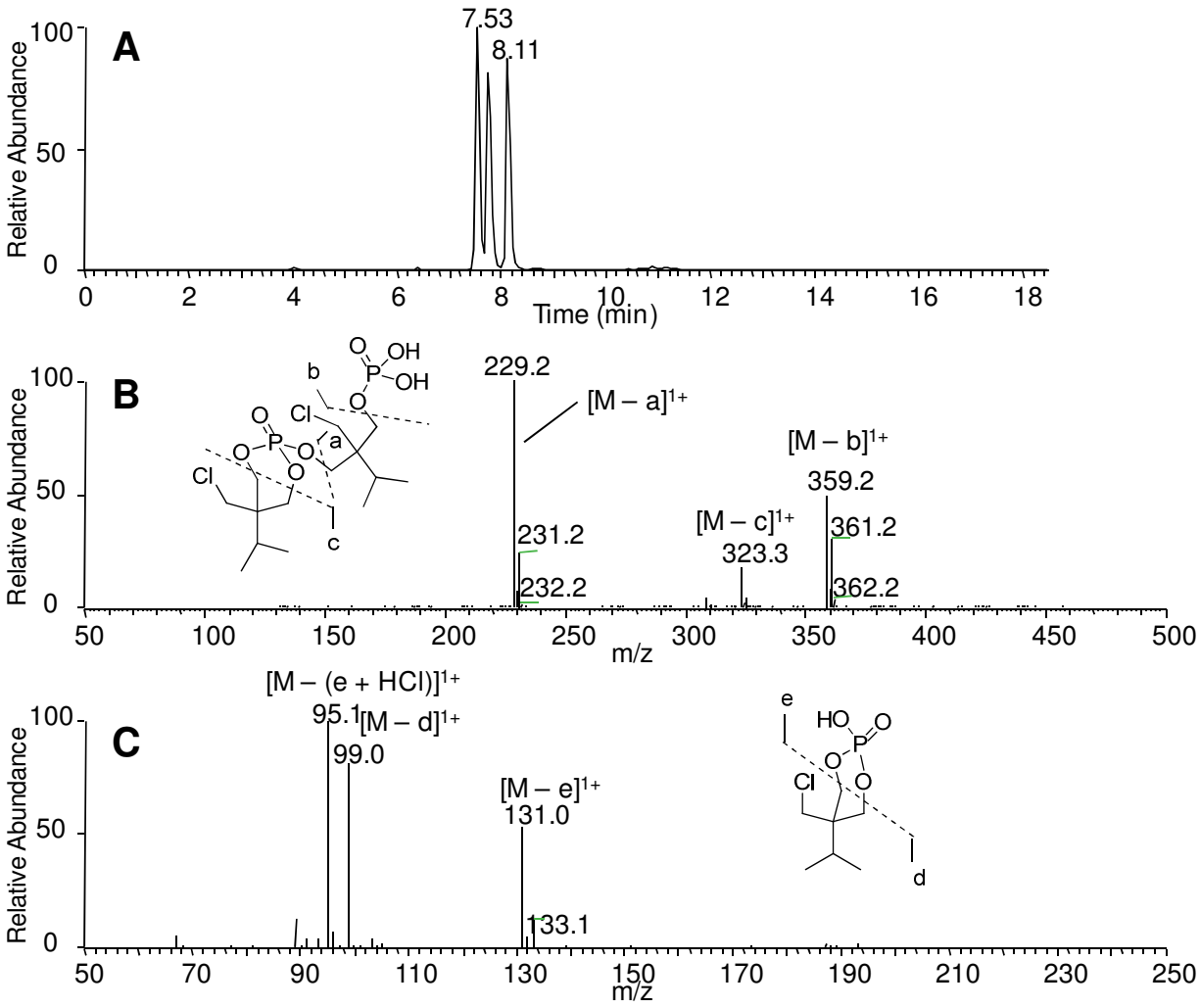


Figure S-2. LC-MS extracted ion chromatogram and APCI-MS^n spectra for *m/z* 457 from an attribution study sample synthesized by Route 4 (RDEN0131): A) Extracted ion chromatogram for *m/z* 457, B) MS/MS of *m/z* 457, C) MS^3 spectrum for the *m/z* 457 → 229 $[M - a]^{1+}$ product ion.