analytical chemistry

Article

Cite This: Anal. Chem. 2019, 91, 15130−15137

pubs.acs.org/ac

Rhapso: Automatic Stitching of Mass Segments from Fourier Transform Ion Cyclotron Resonance Mass Spectra

Remy Gavard,[†] Diana Catalina Palacio Lozano,[‡] Alexander Guzman,[¶] David Rossell,[§,∥] Simon E. F. Spencer,[§] and Mark P. Barrow*[,‡]

[†]MAS CDT, University of Warwick, Coventry, CV4 7AL, United Kingdom

[‡]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

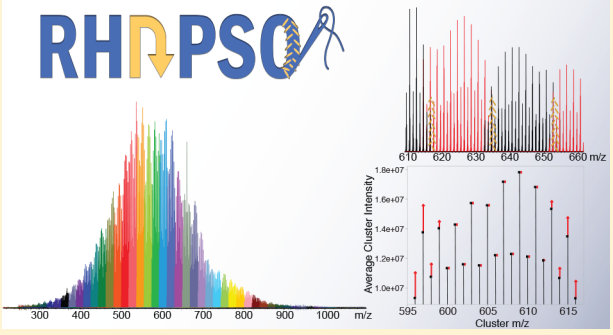
[¶]Instituto Colombiano del Petróleo, Piedecuesta, 681011, Colombia

[§]Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom

[∥]Department of Economics & Business, Universitat Pompeu Fabra, Barcelona, 08005, Spain

Supporting Information

ABSTRACT: Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS) provides the resolution and mass accuracy needed to analyze complex mixtures such as crude oil. When mixtures contain many different components, a competitive effect within the ICR cell takes place that hampers the detection of a potentially large fraction of the components. Recently, a new data collection technique, which consists of acquiring several spectra of small mass ranges and assembling a complete spectrum afterward, enabled the observation of a record number of peaks with greater accuracy compared to broadband methods. There is a need for statistical methods to combine and preprocess segmented acquisition data. A particular challenge of quadrupole isolation is that near the window edges there is a drop in intensity, hampering the stitching of consecutive windows. We developed an algorithm called Rhapso to stitch peak lists corresponding to multiple different m/z regions from crude oil samples. Rhapso corrects potential edge effects to enable the use of smaller windows and reduce the required overlap between windows, corrects mass shifts between windows, and generates a single peak list for the full spectrum. Relative to a stitching performed manually, Rhapso increased the data processing speed and avoided potential human errors, simplifying the subsequent chemical analysis of the sample. Relative to a broadband spectrum, the stitched output showed an over 2-fold increase in assigned peaks and reduced mass error by a factor of 2. Rhapso is expected to enable routine use of this spectral stitching method for ultracomplex samples, giving a more detailed characterization of existing samples and enabling the characterization of samples that were previously too complex to analyze.



Petroleum is one of the most complex mixtures found in nature and can contain hundreds of thousands of unique elemental compositions within a single sample.[1] The study of petroleum composition has become known as "petroleomics".[1−10] Developing a more detailed understanding of petroleum composition in order to address the challenges of producing and refining crude oil has become increasingly important in recent years.[11−14] Fourier transform ion cyclotron resonance mass spectrometry (FTICR MS)[15−20] is a state-of-the-art technique for petroleomics that provides a significant ultrahigh resolving power and mass accuracy to assign elemental compositions of highly complex samples.[21] FTICR cells can hold a maximum of a few million ions, and singly charged ions will be detected if their presence reaches the detectable amount of at least 50 to 100 ions.[22,23] If several thousands of molecular compositions are present, many species can fall below the detection limit.[24] To overcome this problem, we traditionally use signal averaging, summing the data over several scans (usually several hundred).[25] This method is reaching its limit for extremely complex samples, due to the space-charge effects which lowers the isolation dynamic range and mass accuracy.[26]

The space-charge effect can be addressed by segmented acquisition, a method to obtain a full-range FTICR spectrum when the instrumentation was not able to produce a broadband spectrum.[27,28] The spectral stitching method has gained interest recently as the limits of the broadband techniques start to show.[29,30] A quadrupole mass analyzer is used to select ions within a specified m/z range before being passed to the ICR cell, hence reducing the number of different molecular compositions in the cell and helping to get the molecules above the detection threshold. A complete method

Received: August 22, 2019
Accepted: October 30, 2019
Published: October 30, 2019

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

called "selected ion monitoring (SIM) windows" by its authors was established by Southam et al. for biological samples and later used to enhance relative isotopic abundance measurements by Weber et al.[32] SIM was recently improved and made widely available for mass-spectrometry-based metabolomics and lipidomics.[33] Earlier work by Rodgers et al.[34] and Zabrouskov and Senko[35] used stitched spectra after segmented acquisition; however, the data analysis methodology has not been described or made available. Every application of the SIM method reported a higher number of peaks, an increase in the number of peaks assigned, and a higher mass accuracy. Currently, a complete stitching method was only established for metabolomics and lipidomics samples, where the number of peaks observed is around a few thousand and the maximum mass width investigated is m/z 700.[31] In contrast, in complex petroleum samples, the number of peaks in broadband mass spectra can easily reach ten of thousands[30] and spana width of around 1000 m/z. This means that the methods developed for biological samples are not directly applicable to petroleum samples since the windows size must be adapted to the higher molecular density. In addition, because of the higher mass range, it is not realistic to use an overlap of width of 10 m/z between windows. The calibration tools available are different too. The strategy of Southam et al.[31] relies on having a critical number of isolation windows with an internal standard. Petroleomics researchers try to limit the use of internal calibration to avoid doping the sample, but instead use known molecular series as internal calibrants. No methods that are able to perform stitching using peak lists have been publicly described for petroleomics to date. In 2012, Gaspar and Schrader[29] used the commercial software Xcalibur (Thermo Electron, Bremen, Germany) to recreate a full spectrum by adding all the segments acquired, as well as those from the broadband. Recently Krajewski et al.[30] performed the stitching by manually trimming the best width of 20 m/z out of a width of 25 m/z acquisition and ensuring that there was no overlap with the following windows to prevent duplicating peaks.

Southam et al.[31] observed a phenomenon that they called an "edge effect" consisting of a reduction in intensity at the isolation windows' edges compared to what was expected by studying the ratio of two peaks depending of their position across the window. The strategy employed by the authors to account for this effect was to use a large overlap between windows (roughly a width of 10 m/z), so the edge of one window is covered by the central part of the subsequent window (where there is no edge effect). This strategy cannot be practically translated to petroleum samples, as these require substantially smaller windows; increasing their overlap would result in an experiment that could take days to complete.

In this paper we describe a new algorithm called Rhapso to automatically clean the peak lists, correct the edge effect via a convenient statistical model, and stitch the peak list. Rhapso was the name of a nymph in the greek mythology which derives from a greek verb meaning to stitch. Rhapso was recently successfully used to help achieve the highest resolving power and number of unique molecular assignment to date.[36]

METHODOLOGY

The quadrupole was used to only transmit ions within a specified m/z region for detection. This creates an isolation window of a mass spectrum; a partial mass spectrum results, which will be referred to as a "segment". The user keeps the width of the isolation window the same (e.g., spanning a width of 20 m/z) but progressively moves the center of the isolation window to higher m/z (e.g., m/z 261, 279, 297, etc.). In this way, the user acquires a large number of overlapping segments which span the entire m/z range of interest. In Rhapso each segment is trimmed at the high and low m/z ends of the observed signal to enable good subsequent stitching and prevent the inclusion of noise or low quality peaks, producing reduced-width segments. The reduced-width segments can then be combined appropriately, finding suitable regions for overlap, to produce a new mass spectrum.

Sample Preparation. A South American vacuum residue sample obtained using supercritical fluid extraction was used to illustrate our data analysis methodology. The sample has around 90% of its constituents with a boiling temperature less than 720 °C at atmospheric equivalent temperature (AET). High performance liquid chromatography (HPLC) grade toluene (Fisher Scientific, Loughborough, UK) was used to dilute the sample to a concentration of 0.05 mg/mL.

Instrumentation. Mass spectra were acquired using an Apollo II atmospheric pressure photoionization (APPI) source, coupled to a 12 T solariX FTICR mass spectrometer (Bruker Daltonik GmbH, Bremen, Germany). The injection was performed using a flow rate of 500 mL h^{-1}, vaporizer at 350 °C, drying gas at 250 °C, and capillary potential at 1200 V. Potentials of 0.4 V were applied to the front and back trap plates of the ICR cell. For the broadband mass spectrum, a data size of 4 M was used with a detection range of m/z 250−3000, and 100 time-domain transients were coadded. To produce the stitched data, the m/z range (equivalent to 1000 Da) was segmented into 41 windows, each with an m/z width of 20, with each window overlapping the adjacent windows by an m/z width of 2. A quadrupole was used to isolate these narrow m/z ranges, and 50 time-domain transients were coadded.

In order to avoid influencing the peak abundance, the mass envelope, excitation range, magnitude, and ion accumulation time were kept constant. After acquisition, a single zero fill and Sine-Bell apodization were applied before usage of a Fourier transform.

Signal Processing. The FTMS Processing 2.1.0 software was used with an asymmetric apodization ("Kilgour")[37] function for offline phasing of the segments to generate absorption-mode spectra. The spectra were then exported from solariXcontrol to DataAnalysis 4.2 and then internally recalibrated using the HC class with a mass difference of 2.01565 Da. Finally, the peak finder "FTMS" method was used to extract peak information and provide a peak list for each segment and the broadband mass spectrum.

Statistical Processing. Rhapso consists of four steps as depicted in the flowchart in Figure 1.

Step 1: *Removal of Peaks Outside of the Isolation Window.* Peak finding algorithms (e.g., the DataAnalysis 4.2 FTMS peak picking algorithm used in our examples) may identify peaks outside of the target m/z isolation window. In addition, the width of each segment can be different.

We have developed a strategy to obtain reduced-width segments that have a common width and contain the peaks of interest. One option would be to ask the user to input the theoretical m/z range targeted by each window; however, the observed m/z range can differ from the theoretical one due to the precision of isolation of the quadrupole. Instead, we developed a method to detect automatically the m/z range using as little prior information as possible. The method allows

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

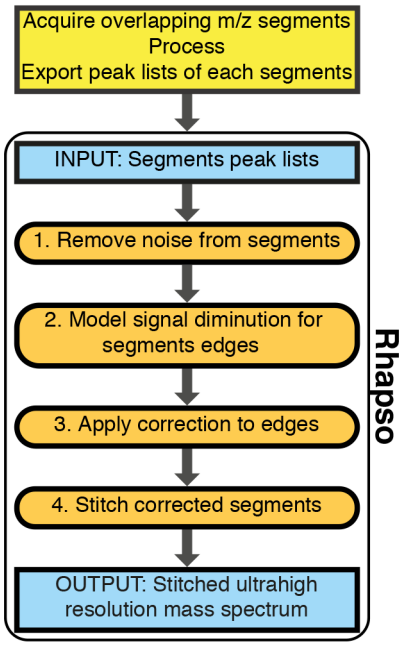


Figure 1. Flowchart representing the four processing steps of Rhapso taking place after acquisition,signal processing and export to peak list for each segment.

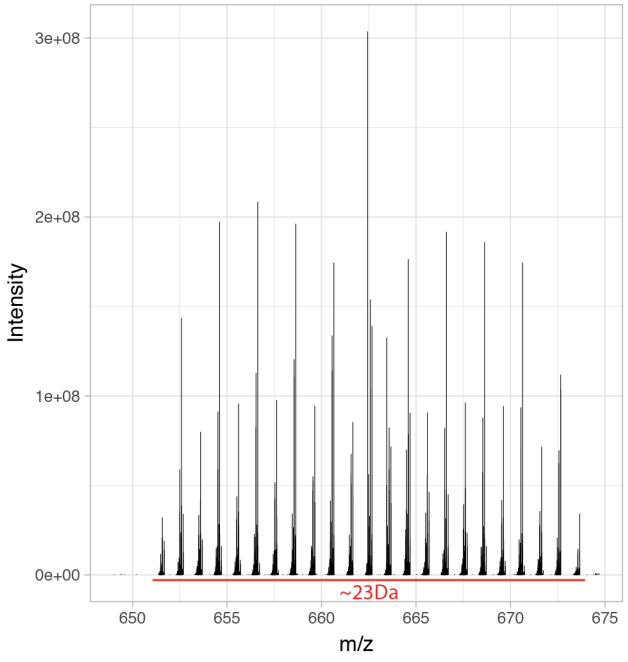


Figure 2. Mass spectrum of a segment between m/z 651 and 674 after FTMS peak picking, before processing and stitching with Rhapso.

for a maximum overlap between consecutive segments of 50% the size of the segments; e.g., for a width of 20 m/z, the maximum overlap is a width of 10 m/z.

Each segment was acquired aiming for a theoretical m/z width, which we denote by W and was kept constant for all spectra. Let r be the number of segments and n_j be the number of peaks in segment $j = 1, ..., r$. Define $M_{i,j}$ as the m/z value of peak $i = 1, ..., n_j$ in segment j and $I_{i,j}$ as the intensity value of peak $i = 1, ..., n_j$ in segment j. We define $S_j = min_i M_{i,j}$ to be the smallest m/z $M_{i,j}$ of segment j and H_j the highest m/z $M_{i,j}$ of segment j. Further, let E_j be the integer m/z at the center of the segment as specified by the user. The observed width of the signal in each segment is often larger (up to 20%) than W).

Our goal is to stitch segments using as wide a range as possible, subject to the measured intensities being high enough to ensure that measurements are accurate. Specifically, we seek the most suitable width $W + x$ to use for the stitching where x is a width adjustment for all segments to be determined as described below. Crude oil molecules are detected in clusters of peaks occurring every integer; consequently, we will investigate segments with widths $W + x$ for integer $x = 0, 1, ..., \lceil 0.2W \rceil$. The cleaning procedure described below is repeated for all the values of x. For each $(W + x)$ and within each segment j we look for the m/z value which maximizes the sum of the intensities in a window of width $(W + x)$. Hence $d_j = \arg\max_d \sum_{i \in S(d)} I_{i,j}$ where $S(d) = \{i : d \le M_{i,j} \le d + (W + x)\}$ within an m/z interval $[d, d + (W + x)]$. The range of d considered for segment j is given by the interval $d \in [E_j - (W + x) + C_j, E_j + C_j]$, with C_j being the decimal which needs to be added to the integer m/z in order to ensure that a cluster of peaks does not get split. If the peak density is too high to determine a space between them, a region with low intensity peaks will be selected. In order to calculate C_j we search for the largest 10 gaps between peaks within $[E_j - (W + x), E_j + (W + x)]$, and for those 10 gaps we calculate the decimal places of $(M_{i+1,j} + M_{i,j})/2$ to identify how far the centers of the gaps are from the integers. If the standard deviation of the decimals of those 10 values is under 0.1, then C_j is the average of those decimals; otherwise $C_j = 0$.

Step 2: *Estimate the Edge Effect.* To investigate and measure systematic intensity decreases at the segment edges, we combined the data from all the cleaned segments by shifting them on a common new scale. This is because individual segments display natural variability due to the chemistry of the crude oils which can be mistaken as an intensity drop if located toward the edges. In contrast, by stacking all segments one can estimate common patterns in intensity drops. The shifted m/z value for peak i in spectrum j is defined as $Z_{i,j} = M_{i,j} - S_j$. That is, the shifted m/z values range is $Z_{i,j} \in [0, W + x]$. Now we divide the $Z_{i,j}$ into k bins 1 m/z wide and let n_k denote the number of peaks for bin k and $Y_k = \log\left(\frac{1}{n_k}\sum_{i=1}^{n_k} I_{i,j}\right)$ for bin k and $X_k = min_{i,j \in k}(Z_{i,j})$ denote the floor m/z $Z_{i,j}$ of each bin. We model the log of the mean intensity of each bin using a piecewise linear model where a and b define the change points at which intensity starts to drop near the edges. That is,

$$Y_k = \beta_0 + \beta_1[(X_k - a) \times \mathbb{I}(X_k < a)] + \beta_2[(X_k - b) \times \mathbb{I}(X_k > b)] + e_k \quad (1)$$

for $k = 1, ..., W + x$ where e_k is an error term and $\mathbb{I}$ the indicator function. The model is fitted by least-squares, which is finding $(a, b, \beta_0, \beta_1, \beta_2)$ that minimizes the mean square residuals (MSR). Specifically, given (a, b) the optimal $(\beta_0, \beta_1, \beta_2)$ can be found by ordinary linear regression. Hence, it suffices to consider a grid of (a, b) values for $a \in \{0, ..., \lceil 0.2(W + x) \rceil\}$ and $b \in \{(W + x) - \lceil 0.2(W + x) \rceil, ..., (W + x)\}$, to find the MSR associated with the optimal $(\beta_0, \beta_1, \beta_2)$ and choose the (a, b) attaining the smallest MSR overall. This step was performed for each of the window sizes $W + x$ investigated. The smallest MSR obtained from all the different window sizes

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

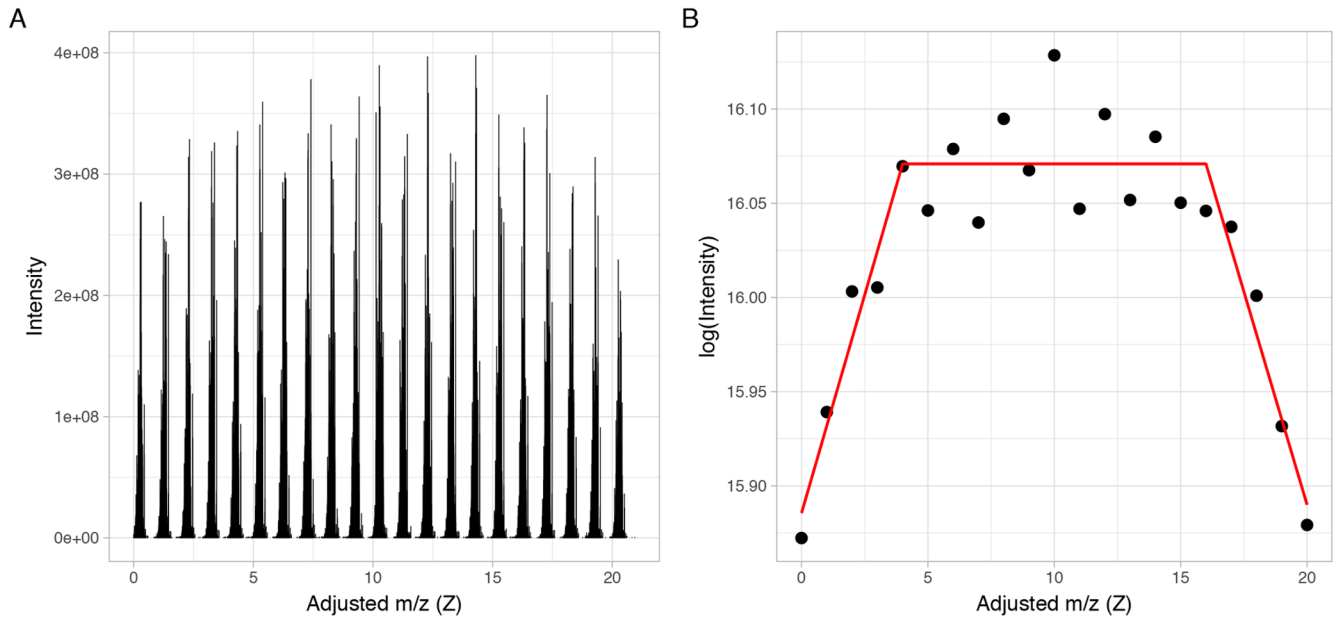


Figure 3. (A) Overlapped log-intensities across all windows in the shifted m/z scale $Z_{i,j}$ and (B) plot of the average intensity of each bin with an m/z width of 1.

is isolated, determining the value of (x, a, b) to be used in the rest of the processing.

Step 3: *Correct Window Edge Effects.* We used the piecewise linear model in eq 1 to correct intensity drops at the segments' edges of each individual segment. Because both the overall intensity and the magnitude of the intensity vary across segments, we estimate β_0, β_1, β_2 separately for each window j. We calculate β_1 and β_2 for each segment and use the associated segment center to fit a locally estimated scatterplot smoothing (LOESS) model[38] for each β. The models are then used to calculate a smoothed β_1 and β_2. This is done to avoid being affected by outlier peaks, especially in the presence of a low number of peaks at the extremities of the window, which can result in an inappropriate value. We denote these new estimated coefficients as $\widehat{\beta_{0j}}$, $\widehat{\beta_{1j}}$, $\widehat{\beta_{2j}}$. The corrected intensities $\widehat{I_{i,j}}$ are obtained as

$$\log(\widehat{I_{i,j}}) = \begin{cases} \log(I_{i,j}) + \max\{\beta_{1j}, 0\}((S_j + a) - M_{i,j}) & \text{if } M_{i,j} < S_j + a \\ \log(I_{i,j}) + \max\{|\beta_{2j}|, 0\}((H_j - b) - M_{i,j}) & \text{if } M_{i,j} > H_j - b \end{cases} \quad (2)$$

where S_j and H_j are the lowest and highest m/z values in segment j.

Step 4: *Stitching.* We have an overlap between segments j and $j + 1$ and we must choose from which window to take peaks. We wish to determine the best m/z to change from segment j to $j + 1$. We isolate the peaks from both spectra with an m/z within the interval $[S_{j+1}, H_j]$ and form a new set $u(j) = \{M_{i,j}: M_{i,j} \geq H_j\} \cup \{M_{i,j+1}: M_{i,j+1} \leq S_{j+1}\}$. Let $\widehat{M_{i,j}}$ be the elements of $u(j)$. We search for the top k largest gaps $\arg\max_i M_{i+1,j} - M_{i,j}$ between peaks in the overlap region $u(j)$. Define P_1, ..., P_k to be the midpoints of these k gaps. In addition, we define $P_0 = (S_{j+1} + H_j)/2$ to be the center of the overlap. The merging point P^* is selected to be the midpoint closest to the center P_0 : $\arg\min_l |P_l - P_0|$. Once all the merged peaks have been identified and the excess peaks removed from each end, the isolation windows are assembled into a unique peak list and exported for further analysis.

RESULTS AND DISCUSSION

The calibrated and phased mass list of each segment was exported using DataAnalysis 4.2 as text files and processed with Rhapso. Rhapso has been implemented using a Shiny web interface.

At this stage, the peak list of each isolation window can contain peaks from outside the targetted isolation due to the noise. The lower the signal-to-noise ratio (S/N) used for the peak picking, the more noise that will be included. As seen in Figure 2, different segment widths can be considered. For instance, the window illustrated in Figure 2 has an observed width of 23 m/z, larger than the theoretical width W of 20 m/z. It is in our best interest to retain as much of segments as possible to be efficient and use fewer segments. In our application, we have explored using widths of 20, 21, 22, and 23 m/z. The m/z range to include for each of the widths considered was determined as described in Step 1. In the subsequent steps we explore each of them and determine which one is the best. As explained in step 2 of the method, we have modeled and corrected a decrease in intensity at the edges of the spectra caused by the isolation by the quadrupole. This phenomenon was described by Southam et al. and called an "edge effect". Instead of using a large overlap and deleting the edges, which would require more segments and in consequence a much longer acquisition time, we decided to

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

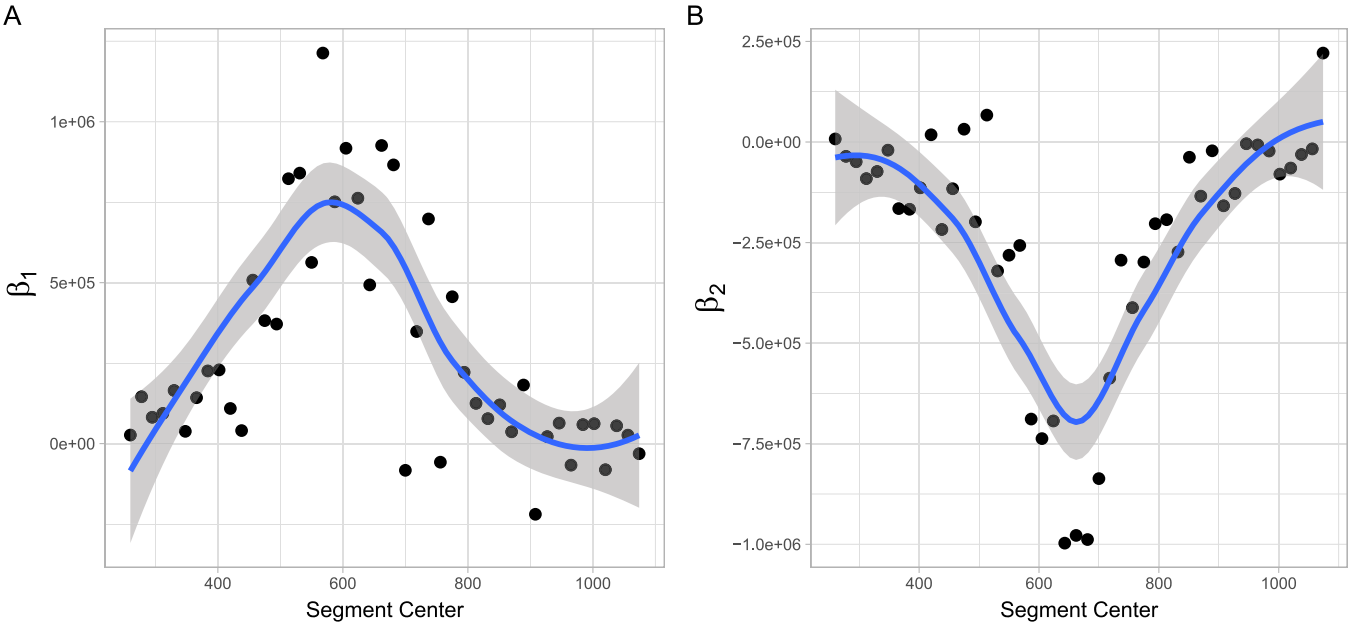


Figure 4. LOESS model estimate of (A) $\widehat{\beta}_{1j}$ and (B) $\widehat{\beta}_{2j}$ for a width $W = 21$, as a function of the segment center.

apply a correction of the intensity. In order to calculate a piecewise model of the intensity diminution at the edges, we had to find where the decrease in intensity occurs. This is specific to the type of sample analyzed, the molecular abundance, the size of the windows, and the instrument. Because of its chemical composition, crude oil displays natural undulations that need to be preserved. In Figure 3 A, we subtract the minimum m/z of each window to all the peaks present in the window and calculated the log of the mean intensities in each bin of width 1 m/z. This allows us to go beyond the natural undulations and to display a clear pattern of intensity diminutions at the edges of the windows. By looking for the minimum mean squared residuals of a piecewise model, we determine the optimum break points, where an intensity correction is needed (Figure 3 B).

Using a grid search for the best breakpoints a and b, the MSR for the model was used to determine the best values for each width considered. Finally, the width with the model yielding the minimum MSR determines the width used in the subsequent steps. After the optimal width W was determined, the β coefficients were calculated for each segment (Figure 4) and a LOESS model was fitted. Figure 4 was created without any log transformation on the intensity to highlight the similarities of the distribution of the coefficients with the intensity profile of the final stitched spectrum. It becomes clear that using the same β for all segments would not be a good fit and proves that the intensity drop at the edges is more pronounced when the intensity increases.

The correction was applied using a β coefficient calculated on the log(intensity) scale and applied to log(intensity) data before being transformed back to the original scale.

In order to assess the quality of the intensity correction performed during step 3, we looked at the mean intensity of each window of width 1 m/z. Figure 5 represents these averaged corrected and uncorrected intensities for each peak cluster. The dots represent the mean intensity without correction, and the arrows point where the mean intensity is after correction. We notice that the correction applied helps restore the natural undulation profile. In accordance to the distribution observed in Figure 4, the correction was most visible in the most intense segments while, at the edges (low and high m/z), the correction was minimal, and sometimes not necessary.

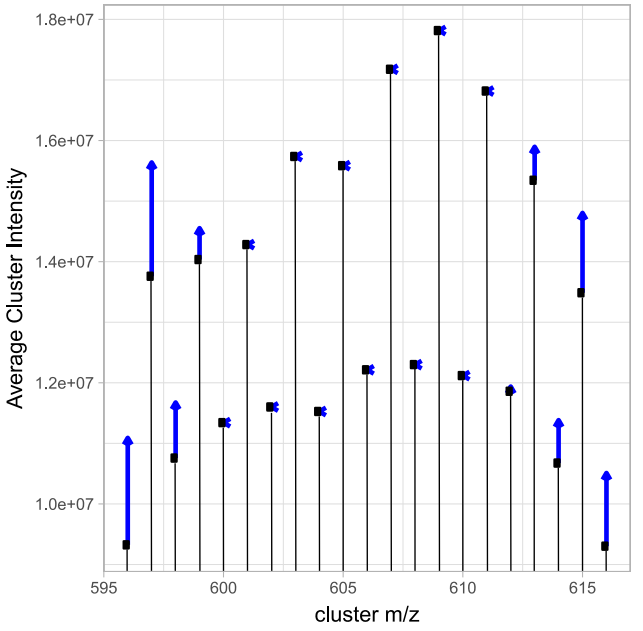


Figure 5. Mean intensity of each window with an m/z width of 1. The arrows indicate the changes to mean intensity of each peak cluster after correction.

After the intensity correction was applied, the merge between each segment was performed. As described in the methods and illustrated in Figure 6A, the overlap region was isolated (illustrated by the two vertical lines). Since we know that the quality of the peaks deteriorates at edges and we want to prevent either duplicating or losing any peaks, the best place to switch from one segment to the other will be in the center of the overlap region and between clusters.

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

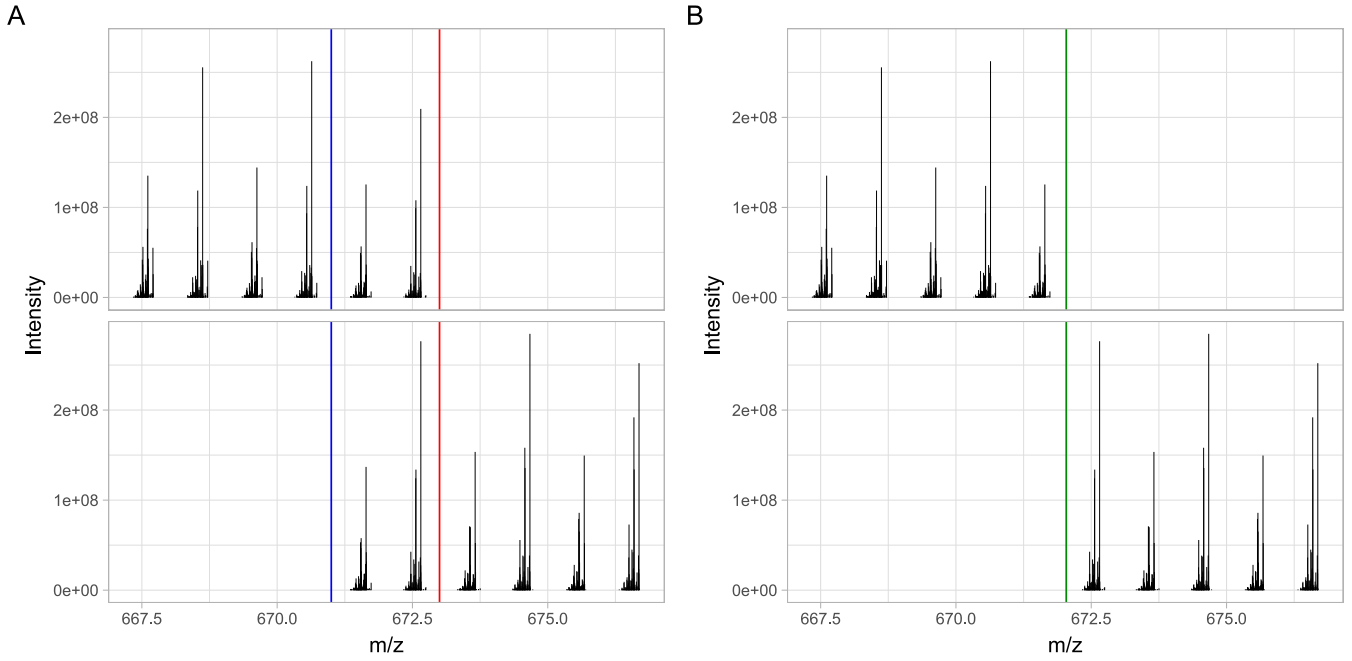


Figure 6. Illustration of the isolation of the overlap region between two segments (A) before and (B) after removal of the unnecessary peaks at the extremities, following automated calculation of the overlap position for segments.

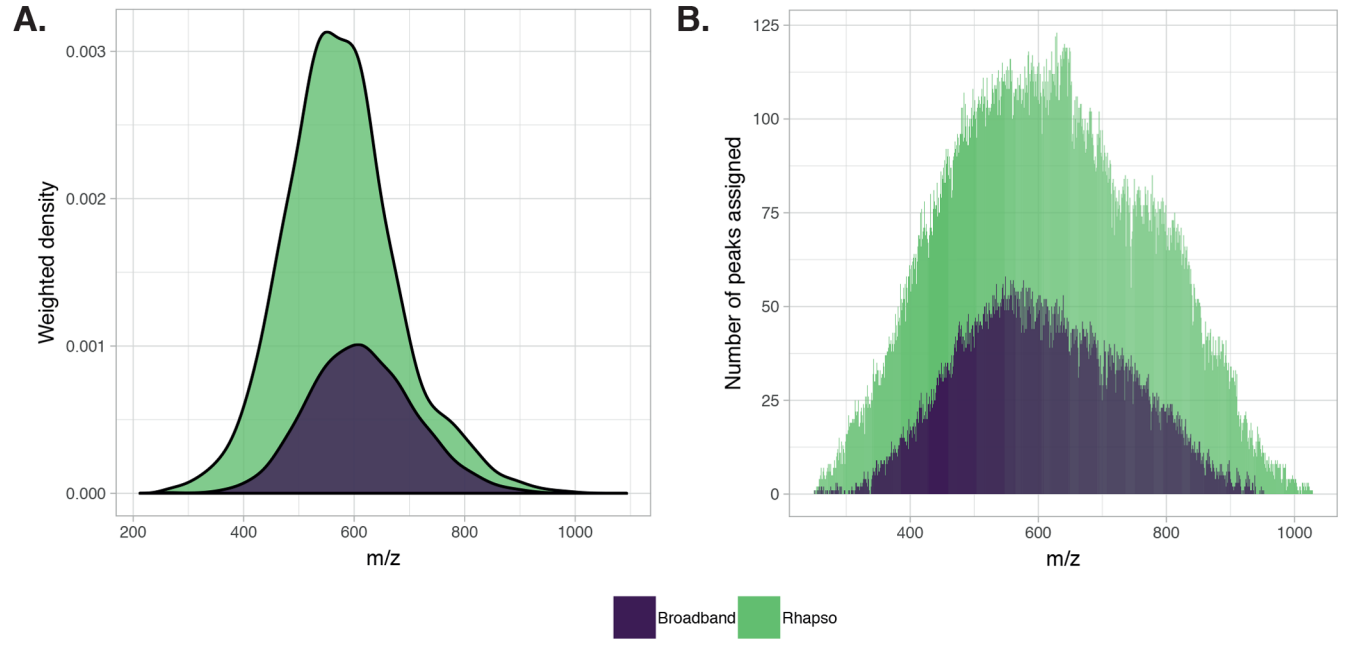


Figure 7. (A) Density plot based on all the m/z measured in both broadband and stitching mode. The plot shows the broader peak distribution in stitching mode due to the increased number of peaks in the low intensity regions. (B) Comparison of the number of assignments per m/z width of 1 using both techniques.

In Figure 6B, the green central vertical line illustrates the m/z where Rhapso found it would be the best place to perform the stitch.

With thousands of peaks, comparing two mass spectra can be challenging, so, in Figure 7A, we have represented a density plot using all the m/z measured in each data set. The density was weighted by the intensity of each m/z. The figure clearly illustrate the higher complexity of the mass spectrum, the higher intensity, and also the broader distribution of the spectrum when using a spectral stitching method.

In Figure 7B, the m/z of each assigned peak was reduced to an integer, and we counted the number of peaks of each integer. The results were plotted as a bar plot and demonstrate the increase in number of peaks assigned after using Rhapso to stitch the segments. It is worth noting the similarity of the distribution with the density plot in Figure 7.

Figure 8 shows the evolution of the main molecular classes. The bar plot clearly demonstrates an increase in the number of peaks assigned for each class. There is a 2- to 3-fold increase in

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

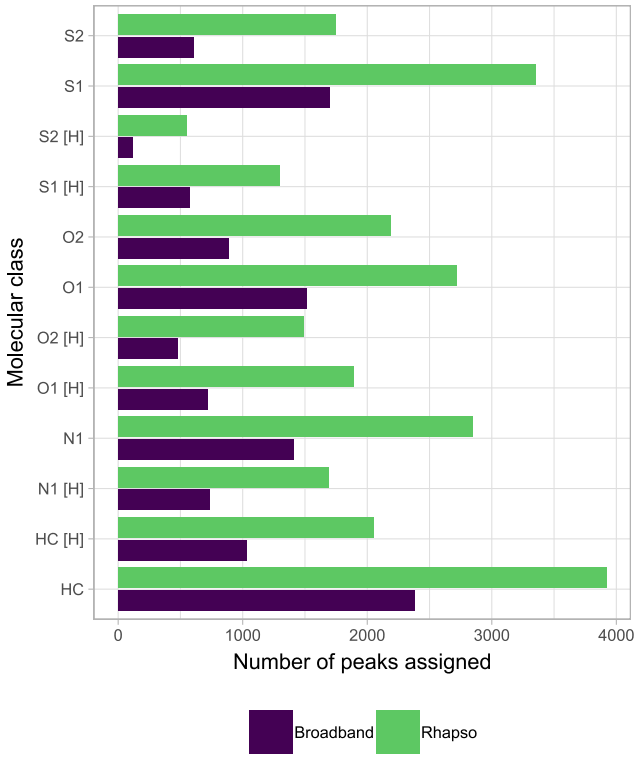


Figure 8. Number of peaks assigned for the prevalent molecular classes with the broadband mass spectrum (purple) and the mass spectrum obtained using segments stitched with Rhapso (green).

the number of peaks assigned in each class which is consistent with the higher number of peaks assigned.

While a higher number of peaks were assigned, this increase did not result in a higher mass error. Figure 9 illustrates how the mass error compares between broadband and stitching. Figure 9A shows the root-mean-square error along the m/z axis. The RMS error was calculated by pooling the mass error, in ppm, for each ppm width of 1 across the range. The data set processed with Rhapso has a systematically lower error and a wider m/z range as already illustrated before. Figure 9B illustrates the distribution of the mass error in ppm observed for all the peaks assigned. The horizontal lines respectively mark the 5%, 25%, 50%, 75%, and 95% quantiles within each violin plot. It is worth noting the much narrower distribution and median closer to 0 with the stitching method.

As presented by Palacio Lozano et al.,[36] Rhapso enabled for this vacuum residue sample a sharp increase in the number of peaks assigned (17k vs 50k) and led to a 2-fold decrease in the RMS mass error. An increase in the number of classes, highest DBE, and highest carbon number were also registered. We have also noted a sharp increase in the number of isotopic peaks assigned.

CONCLUSION

The acquisition process is currently very time-consuming but is expected to be automated in the future. Rhapso performs the spectral stitching of any complex spectrum acquired using selected ion monitoring windows within a few minutes. The method, implemented within a Shiny interface, allows the user to visualize each segment all along as well as check each step of the processing. Rhapso also allows analysts to speed up the acquisition process by permitting the reduction of the overlap between each window thanks to the intensity correction method. It also preserves the natural undulations of the spectra, characteristic of crude oils. The method also demonstrates that the correction allows the peaks' distribution to be closer to the ones observed in broadband mode. The spectral stitching method is crucial to increase the number of peaks observed and lower the mass error using existing instrumentation available to the users. However, any invest-

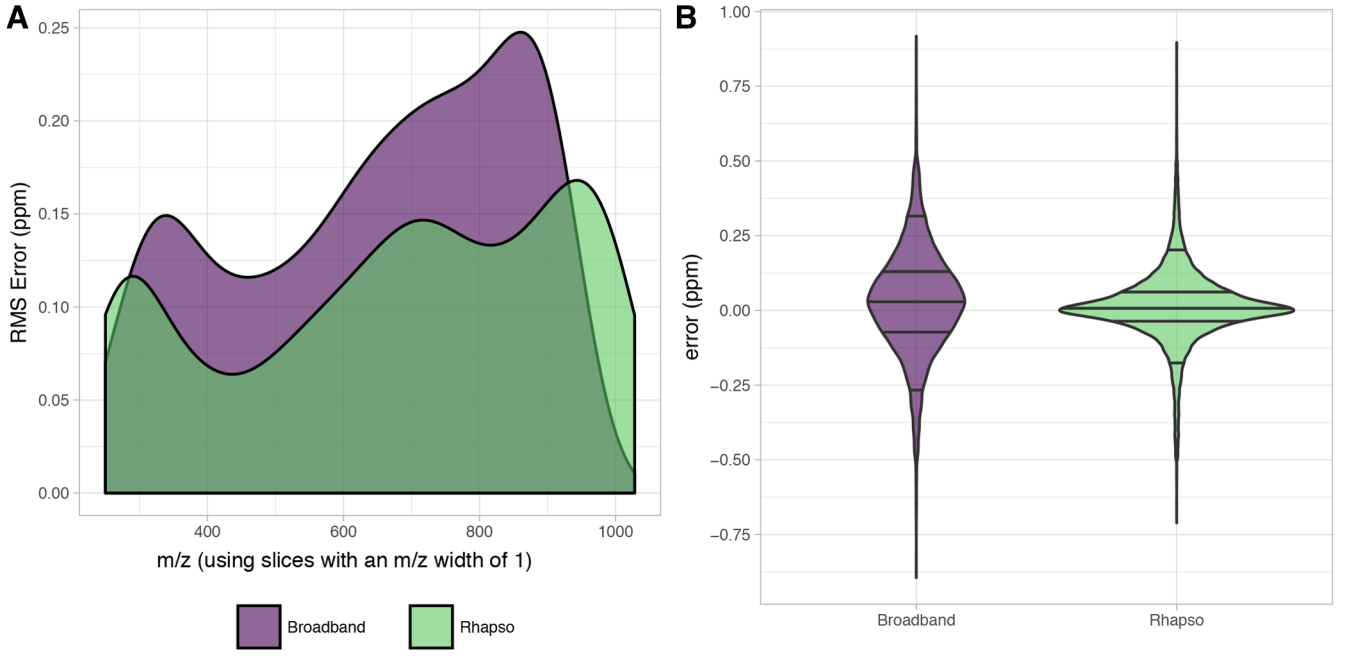


Figure 9. Root mean square error of the assigned peaks for each m/z width of 1 of the mass spectrum as (A) a density plot over the m/z range and as (B) a violin plot.

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137

ment in more expensive equipment such as a more powerful magnet or more advanced FTICR spectrometer will also lead to further improvements. The results demonstrated not only a net increase in the number of peaks assigned but also an increase in the quality of those assignments. While the mass spectrum maintained a similar distribution, we showed there was an increase of peak density in the lower intensity regions. It is expected that this algorithm will enable a more extensive use of this technique, by relieving the user of a highly time-consuming step.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.9b03846.

Figure S1: Broadband mass spectrum of the South American vacuum residue sample, acquired using positive-ion APPI coupled to a 12 T FTICR mass spectrometer. Figure S2: Double bond equivalents (DBE) vs carbon number plots for two heteroatom classes, comparing the traditional broadband experiment and the use of Rhapso. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: M.P.Barrow@warwick.ac.uk.

ORCID

Remy Gavard: 0000-0001-5899-3058
Diana Catalina Palacio Lozano: 0000-0001-5315-5792
Mark P. Barrow: 0000-0002-6474-5357

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

R.G. thanks EPSRC for a PhD studentship through the EPSRC Centre for Doctoral Training in Molecular Analytical Science, Grant Number EP/L015307/1. M.P.B. thank the Newton Fund award (Reference Number 275910721), Research Agreement No. 5211770 UIS-ICP, and COLCIENCIAS (Project No. FP44842-039-2015) for funding. D.R. was partially supported by Ramon y Cajal Fellowship RYC-2015-18544 from Ministerio de Economia y Competitividad (Government of Spain), Ayudas investigación cientifica Big Data (Fundación BBVA), and Programa Estatal I+D+i (Government of Spain). R.G. also thanks David Stranz (Sierra Analytics) for his valuable contributions.

■ REFERENCES

(1) Barrow, M. P. *Biofuels* **2010**, *1*, 651−655.
(2) Marshall, A. G.; Rodgers, R. P. *Acc. Chem. Res.* **2004**, *37*, 53−59.
(3) Barrow, M. P.; McDonnell, L. A.; Feng, X.; Walker, J.; Derrick, P. J. *Anal. Chem.* **2003**, *75*, 860−866.
(4) Ramírez, C. X.; Torres, J. E.; Palacio Lozano, D. C.; Arenas-Diaz, J. P.; Mejia-Ospino, E.; Kafarov, V.; Guzman, A.; Ancheyta, J. *Energy Fuels* **2017**, *31*, 13353−13363.
(5) Palacio Lozano, D. C.; Orrego-Ruiz, J. A.; Cabanzo Hernández, R.; Guerrero, J. E.; Mejía-Ospino, E. *Fuel* **2017**, *193*, 39−44.
(6) Cho, Y.; Witt, M.; Kim, Y. H.; Kim, S. *Anal. Chem.* **2012**, *84*, 8587−8594.
(7) Smith, E. A.; Lee, Y. J. *Energy Fuels* **2010**, *24*, 5190−5198.
(8) Tessarolo, N. S.; Silva, R. C.; Vanini, G.; Pinho, A.; Romão, W.; de Castro, E. V.; Azevedo, D. A. *Microchem. J.* **2014**, *117*, 68−76.
(9) Smith, D. F.; Rahimi, P.; Teclemariam, A.; Rodgers, R. P.; Marshall, A. G. *Energy Fuels* **2008**, *22*, 3118−3125.
(10) Noestheden, M. R.; Headley, J. V.; Peru, K. M.; Barrow, M. P.; Burton, L. L.; Sakuma, T.; Winkler, P.; Campbell, J. L. *Environ. Sci. Technol.* **2014**, *48*, 10264−10272.
(11) Mullins, O. C., Sheu, E. Y., Hammami, A., Marshall, A. G., Eds. *Asphaltenes, Heavy Oils, and Petroleomics*; Springer New York: New York, NY, 2007.
(12) Marshall, A. G.; Rodgers, R. P. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 18090−18095.
(13) Hsu, C. S.; Hendrickson, C. L.; Rodgers, R. P.; McKenna, A. M.; Marshall, A. G. *J. Mass Spectrom.* **2011**, *46*, 337−343.
(14) Headley, J. V.; Peru, K. M.; Barrow, M. P. *Mass Spectrom. Rev.* **2016**, *35*, 311−328.
(15) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *25*, 282−283.
(16) Comisarow, M. B.; Marshall, A. G. *Can. J. Chem.* **1974**, *52*, 1997−1999.
(17) Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* **1974**, *26*, 489−490.
(18) Amster, I. J. *J. Mass Spectrom.* **1996**, *31*, 1325−1337.
(19) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1−35.
(20) Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. *Analyst* **2005**, *130*, 18.
(21) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 3985−3990.
(22) Nikolaev, E. N.; Vladimirov, G.; Boldin, I. A. *Influences of non-neutral plasma effects on analytical characteristics of the top instruments in mass spectrometry for biological research*. American Institute of Physics Conference Series. 2013; pp 281−290.
(23) Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N. *Mass Spectrom. Rev.* **2016**, *35*, 219−258.
(24) Limbach, P. A.; Grosshans, P. B.; Marshall, A. G. *Anal. Chem.* **1993**, *65*, 135−140.
(25) Purcell, J. M.; Merdrignac, I.; Rodgers, R. P.; Marshall, A. G.; Gauthier, T.; Guibard, I. *Energy Fuels* **2010**, *24*, 2257−2265.
(26) Zhang, L.-K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. *Mass Spectrom. Rev.* **2005**, *24*, 286−309.
(27) Guan, S.; Marshall, A. G.; Scheppele, S. E. *Anal. Chem.* **1996**, *68*, 46−71.
(28) Senko, M. W.; Hendrickson, C. L.; Emmett, M. R.; Shi, S. D.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 970−976.
(29) Gaspar, A.; Schrader, W. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1047−1052.
(30) Krajewski, L. C.; Rodgers, R. P.; Marshall, A. G. *Anal. Chem.* **2017**, *89*, 11318−11324.
(31) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. *Anal. Chem.* **2007**, *79*, 4595−4602.
(32) Weber, R. J. M.; Southam, A. D.; Sommer, U.; Viant, M. R. *Anal. Chem.* **2011**, *83*, 3737−3743.
(33) Southam, A. D.; Weber, R. J.; Engel, J.; Jones, M. R.; Viant, M. R. *Nat. Protoc.* **2017**, *12*, 310−328.
(34) Rodgers, R. P.; Hughey, C. A.; Marshall, A. G. *Past, Present, and Future of Environmental Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*. 2002.
(35) Zabrouskov, V.; Senko, M. *Direct Analysis of the Polar Fraction of Heavy Petroleum Crude Oil using a Linear Ion Trap/FTICR Hybrid Mass Spectrometer*. 2005.
(36) Palacio Lozano, D. C.; Gavard, R.; Arenas-Diaz, J. P.; Thomas, M. J.; Stranz, D. D.; Mejía-Ospino, E.; Guzman, A.; Spencer, S. E. F.; Rossell, D.; Barrow, M. P. *Chem. Sci.* **2019**, *10*, 6966−6978.
(37) Kilgour, D. P.; Van Orden, S. L. *Rapid Commun. Mass Spectrom.* **2015**, *29*, 1009−1018.
(38) Cleveland, W. S.; Grosse, E.; Shyu, W. Local regression models. *Statistical models in S*; Chambers, J. M., Hastie, T. J., Eds.; Chapman & Hall: 1992; pp 309−376.

DOI: 10.1021/acs.analchem.9b03846
Anal. Chem. 2019, 91, 15130−15137