34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Visualization of High Resolution Spatial Mass Spectrometric Data during Acquisition*

Mathew Thomas, Brandi S. Heath, Julia Laskin, Dongsheng Li, Ellen Liu, Katrina Hui, Andrew P. Kuprat, Kerstin Kleese van Dam, James P. Carson, *Member, IEEE*

***Abstract*— Mass Spectrometric Imaging (MSI) allows the generation of 2D ion density maps that help visualize molecules present in sections of tissues and cells. The combination of spatial resolution and mass resolution results in very large and complex data sets. New capabilities are necessary for efficient analysis and interpretation of this data. This work details the development and application of the capability to process, visualize, query, and analyze spatial mass spectrometry data. Applications include the generation of 2D maps for selected spectra, the manipulation of the heat maps, and the identification of spectral peaks. Heat maps are generated by projecting the sum of intensity vs. time spectra of each pixel for selected *m/z* value or range. These capabilities take the form of a new interactive software toolkit, MSI QuickView. This software approach is a significant advance over the previous state-of-the art methods that required the conversion of the RAW data using one software, manual assembly of the data, and visualization in another software.**

I. INTRODUCTION

Nanospray desorption electrospray ionization (nano-DESI) is a technique that allows imaging of completely hydrated biological materials with high spatial resolution and sensitivity [1,2]. This significant improvement in detection efficiency and spatial resolution will facilitate new MSI applications in clinical diagnosis, drug discovery, biochemistry, and molecular biology [2]. Imaging nano-DESI experiments, described elsewhere [1,2], utilized a LTQ/Orbitrap instrument and the signal was collected using Xcalibur (Thermo Scientific, Waltham, MA). Xcalibur is a Windows-based user interface that integrates instrument setup, acquisition and data processing for single lines of data from Thermo Scientific mass spectrometers. The typical work flow during data acquisition and processing is illustrated in Fig. 1a. Acquiring MSI data allows the creation of ion density maps for each *m/z* value or signal collected by Xcalibur [3]. However, a limitation of this approach is that data conversion using FireflyTM software (Prosolia Inc.) is limited to 400 scans/ line. In addition, for individual *m/z* values that cannot be processed using Firefly, the line scans have to be manually exported and visualized using Origin Pro 8.5 (OriginLab Corporation, Northampton, MA) [1,2]. Thousands of intensity values are obtained for each particular mass to charge ratio. Due to the large amount of data, it becomes increasingly difficult and time-consuming to manually extract meaningful results from MSI experimentation. Another challenge is that during the Fourier Transform Ion Cyclotron Resonance Mass Spectroscopy (FTMS) mode of acquisition, which is high resolution mass acquisition, Xcalibur does not save the intensity values for each *m/z* at that resolution but instead selects only certain *m/z* values to make the file size manageable and less computationally intensive. Hence there is no certainty on which *m/z* values and how many *m/z* values will be present in each line of data.

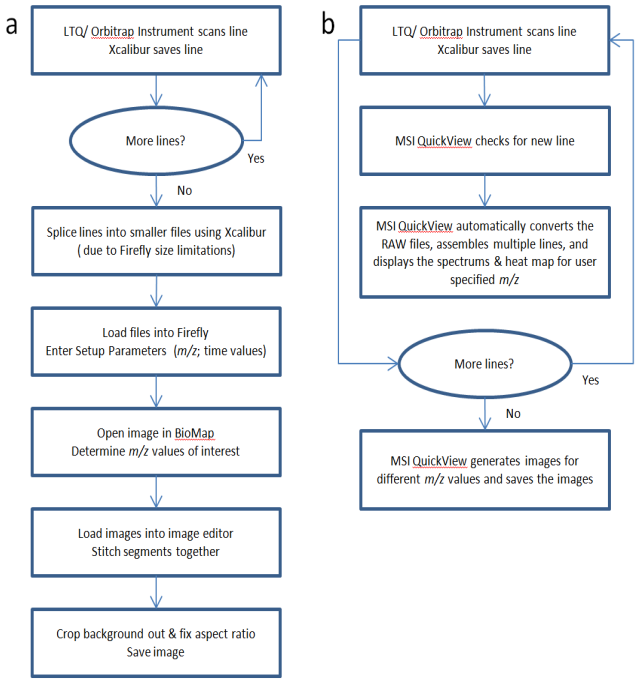

Figure 1. Flowchart depicting (a) the previous workflow without MSI QuickView and (b) the new approach with MSI QuickView.

This paper presents a new approach for handling the data obtained from MSI including FTMS data. The primary objective of this approach is to be able to quickly visualize the data locally over the whole image during image acquisition itself and provide immediate feedback to the user. The method is implemented as a new software named MSI QuickView. The typical work flow using this approach is displayed in Fig. 1b.

*Research supported by U.S. Department of Energy under LDRD DE-AC05-76RL01830.

J.P. Carson is with the Pacific Northwest National Laboratory, Richland, WA 99354 USA (phone: 509-371-6894; fax: 509-371-6946; e-mail: james.carson@pnnl.gov)

M. Thomas, J. Laskin, D. Li, B.S. Heath, K. Hui, E. Liu, A.P Kuprat and K. Kleese van Dam are with the Pacific Northwest National Laboratory, Richland, WA 99354 USA (e-mail: {mathew.thomas, julia.laskin, dongsheng.li, brandi.heath, ellen.liu, katrina.hui, andrew.kuprat, kerstin.kleesevandam}@pnnl.gov)

II. METHOD

A. Software

A software called MSI QuickView which includes a graphical user interface (GUI) has been implemented to speed up experimental feedback via visualization and analysis of MSI data (Fig. 2). As each line of data is acquired, the interface will display the mean intensity vs. *m/z* spectrum, intensity vs. time spectrums for up to 6 different *m/z* values or ranges chosen by the user and heat maps for each line. This assists in validating the usefulness of the particular experiment after scanning the first few lines. In addition, the tool facilitates further processing and analysis of the massive datasets. The user can manually pick different *m/z* values, time ranges, scroll through the spectra for any line in the data without having to load it in manually (Fig. 3), change aspect ratios for the heat maps, and process the heat maps in multiple ways. There is no manipulation of the data required by the user to visualize the data. Example heat maps of a mouse brain section for the same *m/z* value are illustrated in Fig. 2, with the image on the right (bottom) being a smoothed version of that on the left (bottom). MSI QuickView was written in MATLAB® (MATLAB R2011b, Mathworks, Inc., Natick, MA, USA) and compiled using the MATLAB® Compiler™.

B. Interface Workflow

- *Browse for RAW files*

The user selects the first RAW file from a folder to convert. The user can chose to start reading in the data at any line number and also specify the total number of lines to be analyzed.

- *Conversion of the RAW files*

The conversions of the RAW files is accomplished by calling the XConvert program directly from MATLAB, which will convert the Xcalibur RAW data to a CDF (Common Data Format). XConvert is provided by the manufacturers of Xcalibur, Thermo Scientific. XConvert can also convert the RAW files into the more universal TXT or MZXML format. In order to quickly access to the huge datasets, the data from each RAW file is saved as a MAT (MATLAB binary format) file in a temporary folder. This MAT file only contains the necessary values for MSI QuickView, that is, the intensity and *m/z*. The MAT file is deleted once the visualization and analysis of that dataset is completed.

- *Automatic/Manual Mode*

The program can be run in the automatic or manual mode. The automatic mode will wait for each line of MSI data to be saved by Xcalibur and then automatically convert and display the required data. Xcalibur is continuously saving data to the RAW file as each scan in a line is being read. If the automatic mode is selected, the program will not start reading the data for a particular line until Xcalibur starts saving the next line. The maximum of the total time taken for the instrument to go through each line and save the RAW file guides the timing for initiating the reading of the RAW file for the last line. The manual mode requires the user to press the scroll button to display the next line of data. The manual mode thus reduces computer usage during experiments that take a long time to go through each line.

The intensity values for each scan is summed up for the *m/z* value or range specified to obtain the 'Intensity vs. Time' spectrum for each line. The 2D intensity image generated for this gives the heat map for that particular line. This is repeated for each line of data to obtain the heat map for the entire sample. The user can change the aspect ratio of

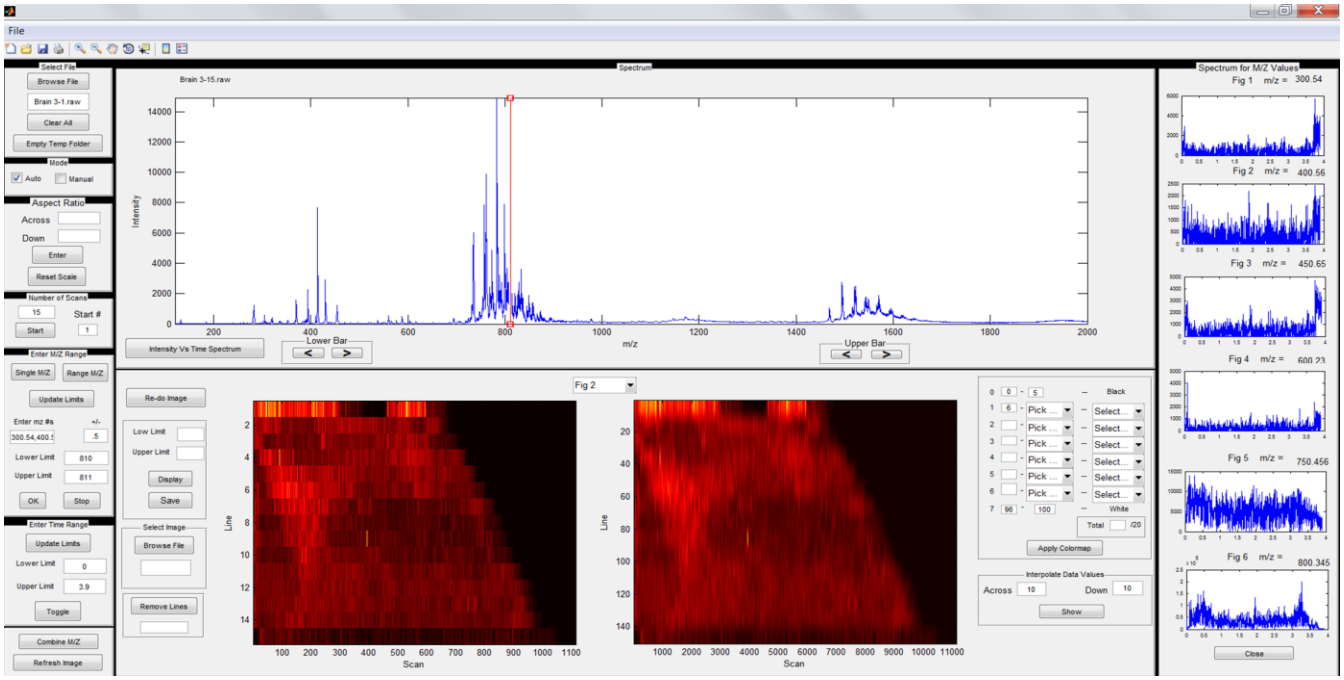

Figure 2. The interactive interface for MSI QuickView

the Intensity vs. Time heat map by entering new values. For generating these heat maps, the user has the option to select either a range of *m/z* value or a single *m/z* value. This can be adjusted by moving the red indicator lines on the spectrum, by using the scroll buttons in the 'Low Bar' box and 'Upper Bar' box or by entering the *m/z* value in the 'Enter *m/z* Range' panel. The user can also enter up to 6 *m/z* values in the 'Enter *m/z* Values' box with an optional threshold value. This will generate Time vs. Abundance spectrum for each *m/z* value and keep updating the plots as each line is read. For a single *m/z* value, the heat map displayed will be the intensity at that *m/z* value for each scan and each line of data. For a range of *m/z* values, this heat map will be the sum of intensities for the range of *m/z* values for each scan and line of data. To view the heat map for a new *m/z* value, the user can select a new range or a single value and use the 'Re-do Image' button to replace the heat map with the new one.

There are often hot spots in the data which take the highest abundance value (have the highest peak) and hence provide no contrast in the other interesting regions of the sample. To avoid this problem, the user can set the upper and lower limits of the intensity values to be displayed in the heat map. Values above or below these limits will be displayed as the upper or lower intensity limit, respectively, and the heat map will be displayed for the new range. The heat map upper and lower limits can be restored to default settings by selecting the 'Refresh Image' button.

The 'Intensity vs. Time Spectrum' button will display the 'Intensity vs. Time' spectrum over the entire range of *m/z* values in the top window and the corresponding 'Intensity vs. *m/z*' spectrum in the bottom window (Fig. 3). The time range can be adjusted by moving the red indicator lines on the spectrum or by entering the new time values in the 'Enter Time Range' window. The user can scroll through the 'Intensity vs. Time' spectrum for different lines, this will automatically display the 'Intensity vs. *m/z*' spectrum for that particular line as well. The user can also scroll through just the 'Intensity vs. *m/z*' spectrum at the bottom window. The 'Intensity vs. *m/z*' spectrum is the mean spectrum of intensity values for the time range specified by the user.

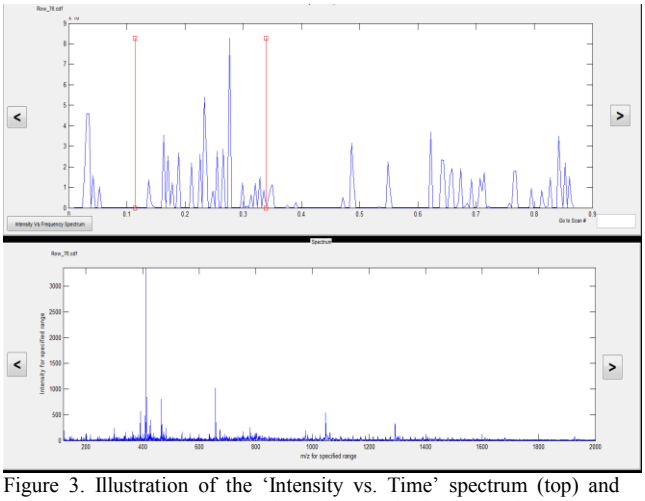

Figure 3. Illustration of the 'Intensity vs. Time' spectrum (top) and 'Intensity vs. *m/z*' spectrum (below) for a particular line of data.

The heat map initially obtained is pixelated and hence does not give the ideal representation of the sample. This is ameliorated by interpolating the data between each scan and line to give it a more smooth appearance. The user can specify the number of interpolated values to add between each scan and line in the 'Interpolated Data Values' window. Other smoothing options available within the toolbar in the program include an averaging filter of user defined radius and a row-wise or column-wise filter. The heat map can be saved using the 'Save Image' button. The output filename will contain information such as folder name, colormap, interpolated values used for smoothing the image, and scaled intensity values. Hence, the user can save multiple heat maps for the same dataset at different settings. There are 2 image windows for displaying heat maps. The user can select which window to have control over by using the drop down box. This will allow the user to simultaneously view the heat map for 2 different *m/z* values or ranges of *m/z* values. Any modifications to image settings will only modify the heat map in the user selected axis. The user can view an enlarged version of the heat map by selecting the 'Enlarge' button in the toolbar window.

In addition to several built-in MATLAB colormaps, the program also has a personal colormap option. The user can specify which range of intensity values get which range of colors. There is also an option to change the color of all intensity values that lie within a range selected by the user, while all other colors remain the same. For this the 'Change Intensity Range' button in the toolbar is used. 'Pick Intensity' button will allow the user to click 2 points on the heat map to get the lower intensity value and higher intensity value. Once you have the range, you can scroll through the 'Inc/Dec Intensity' option to increase or decrease the brightness of just the pixels within that range of intensity.

For the FTMS data, the *m/z* values in each scan and line can be different. MSI QuickView handles this issue for real-time-visualization by reading the *m/z* list from the first line and then building up the *m/z* list as other lines are processed. This allows the user to visualize the FTMS data on the fly. Once the acquisition has been completed, the program corrects the entire dataset to a common *m/z* grid.

C. *Peak Picking*

Due to the large dimensions of data produced during MSI acquisition, manually detecting features of interest from the several thousand *m/z* values is prohibitive. A qualitative understanding of spectral and spatial correlations of different *m/z* values requires the comparison of multiple single-peak images, which becomes time consuming and impractical as number of peaks increases [3]. To address this, a method to perform data processing for observing masses of interest or spatial patterns was needed [4].

A peak in a mass spectrum displays the significant presence of a molecule for that particular *m/z* value or range of *m/z* value [3]. When the user sets the upper or lower *m/z* limit, the program will pick the closest *m/z* values for the upper and lower limit if the exact *m/z* value entered by the user is not present. The important peaks are currently being detected using Decon2LS software and the application of

Principal Component Analysis (PCA). Decon2LS is an open-source software package that can detects useful peaks in MS data [5]. The tool uses an assortment of algorithms for the various parts of the deconvolution process including noise reduction, peak detection, prediction of theoretical isotopic envelope and scoring functions that quantitate the quality of the signature observed in the data [5]. Using Decon2LS, the user can obtain a refined set of peak values by setting parameters that include peak fit type, signal to noise ratio, peak background ratio, etc. This reduces the *m/z* values from several thousands to a few hundreds based on the parameters selected by the user. The number of values can be further reduced by applying PCA, which has previously been used successfully with MSI [3,4,6,7]. PCA is applied on a matrix containing the important *m/z* values for each line to find the *m/z* values that are most prominent for multiple scans and lines. Similar *m/z* values are also removed from the list. This further reduces the list of *m/z* values. Then, a PDF containing the heat maps for a maximum of twenty *m/z* values is generated. The Decon2LS software can be downloaded free of charge at http://omics.pnl.gov/software/.

III. Future Work

The next development steps being pursued include adapting the software to multiple instruments, incorporating cluster-based analysis of the data, adjusting width of data collected to be self-consistent, and multidimensional visualization. For the first of these, MS QuickView is being tested with '.yep' files from Bruker Daltonics systems.

When the user knows very little information about the sample being analyzed, it can be extremely useful to perform and unsupervised analysis to group parts of the image based on their entire spectra [4]. This type of cluster analysis is a very common application on MSI data. The concept is to perform unsupervised clustering of the heat maps for all of the important *m/z* values and identify the different features. A user-supervised K-means clustering method is already implemented. This, however, requires the user to enter the number of features expected for each *m/z* value. Three techniques are currently being implemented for unsupervised classification: PCA, Bayesian minimum mean square error estimation, and minimum correlation function estimation. Fig. 4 shows an example of two images in the same class as identified using a Least Squares approach.

Although the total time for scanning through each line during acquisition is approximately the same, each line can

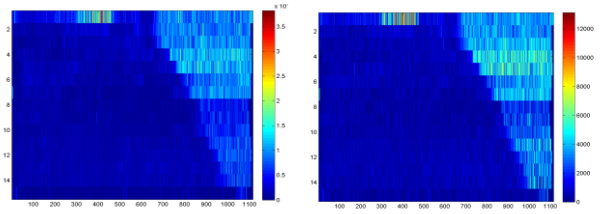

Figure 4. Example of the Least Squares approach for classification. (a) Heat map at *m/z* 283.83 and (b) heat map at *m/z* 413.75 belong to the same class.

have a different number of scans. The total width of the sample will be the same throughout. We are attempting to utilize the time taken for each scan to account for the time difference between each scan and interpolate this difference to make the number of scans for each line the same.

3D visualization of multilayer sectioning data is being developed using ParaView (Kitware Inc., 28 Corporate Drive, Clifton Park, NY 12065, USA). ParaView is freely available and can be setup to employ parallel processing for handling large datasets. This also allows offloading the visualization workload to a remote server dedicated to visualization. Fig. 5 displays an image generated by MSI for a mouse brain tissue slice using ParaView.

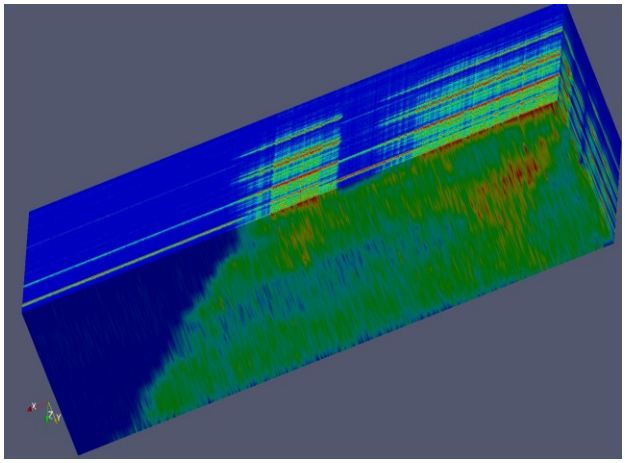

Figure 5. A 3D view generated using ParaView of MSI data. The 'x axis' is the number of scans (1115 voxels wide), the 'y-axis' is the number of lines (15 voxels wide), and the 'z-axis' represents *m/z* values.

References

[1] P.J. Roach, J. Laskin, A. Laskin, "Nanospray Desorption Electrospray Ionization: an Ambient Method for Liquid-Extraction Surface Sampling in Mass Spectrometry." *Analyst*, vol. 135, pp. 2233-2236, 2010.

[2] J. Laskin, B.S. Heath, P.J. Roach, L. Cazares, O.J. Semmes, "Tissue Imaging Using Nanospray Desorption Electrospray Ionization Mass Spectrometry" *Anal Chem*, vol. 84, pp. 141-148, 2012.

[3] G. McCombie, D. Staab, M. Stoeckli, R. Knochenmuss, "Spatial and Spectral Correlations in MALDI Mass Spectrometry Images by Clustering and Multivariate Analysis," *Anal Chem*, vol. 77, pp. 6118–6124, 2005.

[4] E.-R. Muir, I. J. Ndiour, N. A. Le Goasduff, R.A. Moffitt, Y. Liu, M.C. Sullards, A.H. Merrill, JR. , Y. Chen, and M.D. Wang, "Multivariate Analysis of Imaging Mass Spectrometry Data". *Bioinformatics and Bioengineering*, 2007. [Proceedings of the 7th IEEE International Conference on, vol. 1, 14-17th Oct 2007, pp. 472-479].

[5] N. Jaitly, A. Mayampurath, K. Littlefield, J.N. Adkins, G.A. Anderson, et al., "Decon2LS: An open-source software package for automated processing and visualization of high resolution Mass Spectrometry Data," *Anal Chem*, vol. 78, pp. 7397–7409, 2006.

[6] M.S. Wagner, D.G. Castner, "Characterization of Absorbed Protein Films by Time-of-Flight Secondary Ion Mass Spectrometry with Principal Component Analysis," *Langmuir*, vol. 17, pp. 4649–4660, 2001.

[7] R. Van de Plas, F. Ojeda, M. Dewil, L. Van Den Bosch, B. De Moor, E. Waelkens, "Prospective Exploration of Biochemical Tissue Composition via Imaging Mass Spectrometry Guided by Principal Component Analysis," *Pacific Symposium on Biocomputing*, pp. 458–469, 2007.