

## Millions of mass spectral peaks resolved and detected: Now What?

RYAN P RODGERS<sup>1</sup>, CHRISTOPHER L HENDRICKSON<sup>1</sup>,  
TEJA J POTU<sup>1</sup>, CHRISTOPHER H MONTENEGRO<sup>1</sup>,  
ALVARO J TELLO-RODRIGUEZ<sup>1</sup>, PIERRE GIUSTI<sup>2</sup>, BRICE  
BOUYSSIERE<sup>3</sup>, CARLOS AFONSO<sup>2</sup>, CHRISTOPHER P  
RÜGER<sup>4</sup>, MARTHA CHACON<sup>1</sup> AND GERMAIN S  
VALLVERDU<sup>3</sup>

<sup>1</sup>National High Magnetic Field Laboratory

<sup>2</sup>University of Rouen-Normandy

<sup>3</sup>Université de Pau et des Pays de l'Adour

<sup>4</sup>University of Rostock

Realization of the potential molecular information afforded by direct infusion and on-line FT-ICR analyses of complex mixtures (i.e. lipids, metabolites, environmental, petrochemical, pyrolysis oils, aerosols, emerging contaminants) relies on confident mass calibration and subsequent elemental composition assignment. However, mass spectral performance metrics (requisite speed, signal-to-noise (S/N), resolution, mass accuracy) and compositional complexity can vary widely as a function of instrument (most commonly FT- or FT-ICR MS), application, and experiment type. Combined, these factors have created a bottleneck in the overall workflow and hindered the utilization of historic data via data mining and machine learning approaches. Herein, we describe a fully automated data processing workflow based on mass difference analysis (MDA) applicable to a variety of complex matrices.

Mass differences are calculated in a software package (PyC2MC) from high to low  $m/z$  by subtracting the  $m/z$  of all detected peaks from the highest  $m/z$ , and repeated until no peaks remain. The mass difference histogram is plotted as number of occurrences vs. mass difference ( $\Delta m/z$ ), and the center and width at half height ( $\gamma$ ) are determined for all differences. A subset of unambiguous mass difference histogram(s) are used to recalibrate (global) the entire mass spectrum by optimization of the mass calibration coefficients for each mass difference pair in the selected histogram so that the measured  $\Delta m/z$  equals the theoretical  $\Delta m/z$ . After initial global recalibration, these constraints along with accurate mass are used to determine one or more confident molecular formula assignment(s) at low  $m/z$ . These assignments are connected to all other mass peaks by graph theory, utilizing a set of unambiguous mass difference distributions (edges), which enables a subsequent walking calibration and ultimately molecular formula assignment. Finally, statistical verification of molecular formula assignments is facilitated by the relationship between precision, S/N, and resolving power that defines the allowable error for molecular formula assignment based on a user defined confidence. Its application to other mass analyzers (FTMS) will be explored. Work supported by the NSF Divisions of Materials Research and Chemistry through DMR-2128556 and DMR-1644779, FSU, and the State of Florida.