Exploring unidentified peptide sequence data from a circadian rhythm study in *Kalanchoe fedtschenkoi*

Armin Geiger, Paul E. Abraham, Xiaohan Yang, Rongbin Hu, Richard Giannone, and

Daniel Jacobson

Oak Ridge National Laboratory

[geigerag@ornl.gov](mailto:geigerag@ornl.gov)

**Abstract**

Liquid chromatography coupled to two rounds of mass spectrometry (LC-MS/MS) applied in a technique known as `shotgun proteomics', has proven effective as a means to capture a considerable portion of the protein complement of a biological sample. The technique often produces millions of mass spectra per experiment, however, approximately 50-75% of MS2 spectra remain unidentified, even as a good portion of these spectra are of high quality and likely peptide-derived. There are many possible reasons why these spectra may go unassigned including not having good enough database matches for spectra arising from biological phenomena such as unknown post-translational modifications and Single Nucleotide Polymorphisms(SNPs). In addition, problems such as spectral chimerism - a phenomena where the isolation window for a peptide contains more than one distinct peak - is also known to negatively impact spectral library searching. Clustering these high-quality unassigned (HQU) spectra together with their assigned counterparts combined with a spectral purity analysis may yield insight into the origins of these HQU spectra.

This work combines clustering of mass spectra (using 3 distinct algorithms) with a spectral purity analysis. Moreover, the experimental design is leveraged by using the peptide intensities to identify unidentified spectra of possible biological relevance. The method is applied to an LC-MS/MS dataset obtained from a circadian rhythm experiment in the plant species, *K. fedtschenkoi.* This plant is an important model species for the study of Crassulacean Acid Metabolism - a special adaptation of plants that inhabit areas with low water availability. Mining of this untouched proteome resource may yield valuable insight into the proteomic changes that occur during the circadian rhythm of this plant.