**High-throughput Structural Annotation of Massive Natural Products LC–MS/MS Datasets with Next Generation Data Analysis Techniques**

Kyo Bin Kang

*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego*

Liquid Chromatography hyphenated tandem mass spectrometry (LC–MS/MS) is the most frequently used analytical platform for untargeted metabolomics studies involved with natural products, due to its wide coverage and high sensitivity to semi-polar metabolites. However, extraordinary chemical diversity of natural products makes the identification of metabolites to be a significant obstacle in LC–MS-based metabolomics workflows. Here, we introduce a semi-automated workflow that facilitates mapping and identifying massive chemical spaces, based on MS/MS molecular networking enhanced by *in silico* fragmentation and substructure recognition. This workflow classifies a massive LC–MS/MS dataset into molecular families according to their spectral similarity and putatively annotates their chemical scaffolds and substructures even without any spectral library matching. Based on the putative annotations, we are able to characterize, analyze, and visualize the chemical space of natural products, allowing us to digitize and navigate the diversity and distribution of metabolites. We expect this workflow to accelerate metabolomics studies as the Basic Local Alignment Search Tool (BLAST) did for genomics studies.