

Component Elucidator the software for automated analysis of high resolution accurate mass LC-MS datasets in metabolomics

Serhiy Hnatyshyn¹, Tom McClure², Michael Reily¹, Mark Sanders³

¹Bristol-Myers Squibb Co., Rt. 206 & Provinceline Rd, Lawrenceville, NJ, 08543, USA

²Thermo Fisher Scientific, San Jose, CA, USA

³Thermo Fisher Scientific, Somerset, NJ, USA

Successful application of high resolution accurate mass (HRAM) LC-MS analysis in drug discovery requires timely delivery of results to decision makers. While HRAM LC-MS provides excellent sensitivity, accuracy, wide dynamic range and is amenable to high throughput automation, it also produces an enormous amount of information. A typical 10 minute HRAM LC-MS profile of a biological sample may contain over a million data points. The sheer volume of information obtained in a short HRAM LC-MS run completely precludes thorough manual data interpretation. Currently available software solutions are inadequate to meet data processing demands with the constraint of industrial turn-around requirements.

To address this significant gap in the application of HRAM LCMS in drug discovery, Component Elucidator (CE), a collection of data mining algorithms constituting a method for fully automated systematic analysis of high resolution data, was developed using the LTQ-Orbitrap (Thermo Fisher) HRAM LC-MS platform. BMS has formed an alliance with Thermo Fisher Scientific Inc., to package and commercialize CE as a complete data analysis software product. The software includes LC-MS data processing, signal annotation and statistical analysis. The data processing module reduces millions of data points to a few hundred real “components” by 1) eliminating noise, 2) peak filtering and 3) combining information from isotopic peak profiles, adducts, and dimers into a single accurate mass and retention time corresponding to a unique analyte. In the annotation module, the components are assigned molecular identities using various knowledge bases and databases. CE then uses sophisticated alignment procedures to generate an output table of annotated components along with relative abundance in each sample. Finally, in the statistical module, univariate or multivariate statistics can be applied across the samples in a study, to reveal changes and trends that can be correlated with various biological or analytical endpoints.