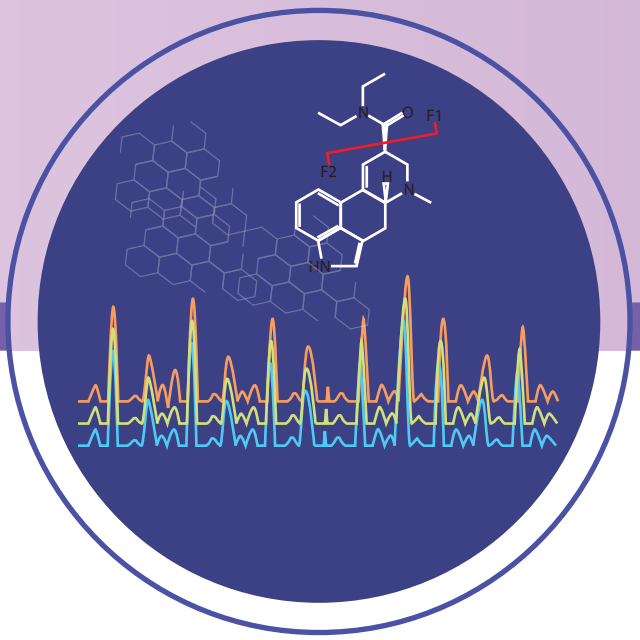
PREMIER Biosoft

SimMet®

High Throughput Bioinformatics Tool for Mass Spectrometry Based Metabolomic Data Pre-Processing and Analysis



SimMet® is a comprehensive software suite for LC-MS data processing, metabolite identification and statistical analysis. It can read raw data from all the major mass spectrometer manufacturing vendors. It provides a highly efficient platform for discovery metabolomics, right from raw data processing to reliable compound identification that are of statistical significance.

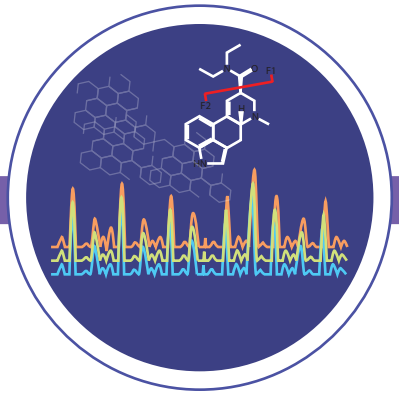
The program accepts raw data in:

- **Vendor-specific native data file formats –** *.fid, *.baf and *.yep (Bruker Daltonics), *.wiff (SCIEX), *.raw (Thermo Scientific™)

- **Standard data file formats namely –** text, MS Excel, mzData, and mzXML

The software enables metabolite identification through searching of a proprietary and carefully curated database which includes 66,512 metabolite species from various biological sources such as Human, Yeast and E. coli. NIST MS/MS database containing 9,390 compounds with 45,298 ions from 234,284 MS/MS spectra is also supported to facilitate experimental validation.

SimMet®



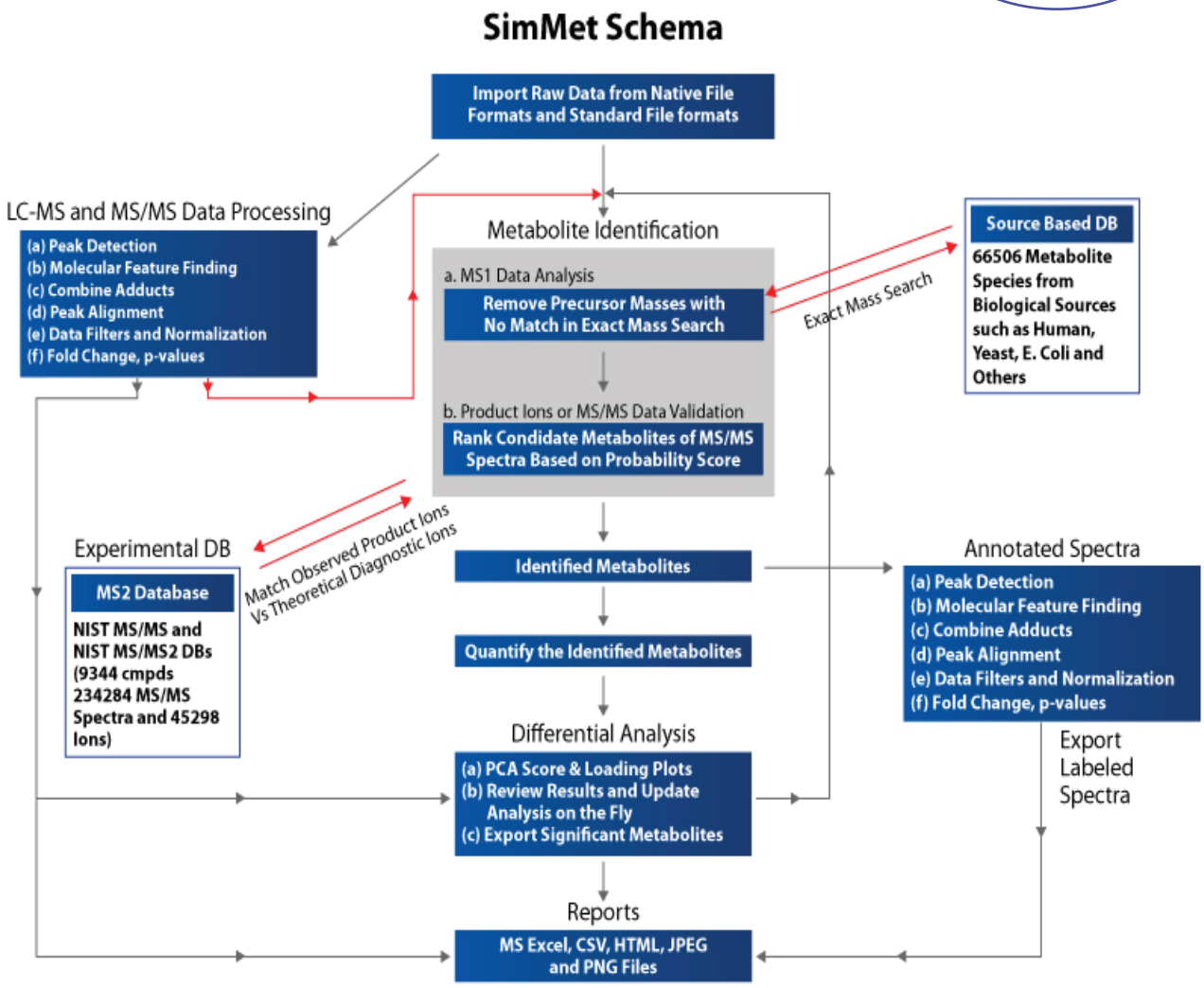


Figure 1: Schematic representation of SimMet software workflow for metabolite profiling using source based databases and experimental databases, LC-MS and MS/MS data processing and further downstream processing using statistical methods using MS and MS/MS data

KEY FEATURES

1. LC-MS Data Processing: Automated LC-MS data processing steps include peak detection, peak picking, molecular feature finding, retention time alignment across experimental LC-MS runs. Choice of scaling and normalization techniques is provided prior to statistical analysis for elimination of unwanted systematic bias, while maintaining genuine biological differences in the observed datasets.

SimMet®

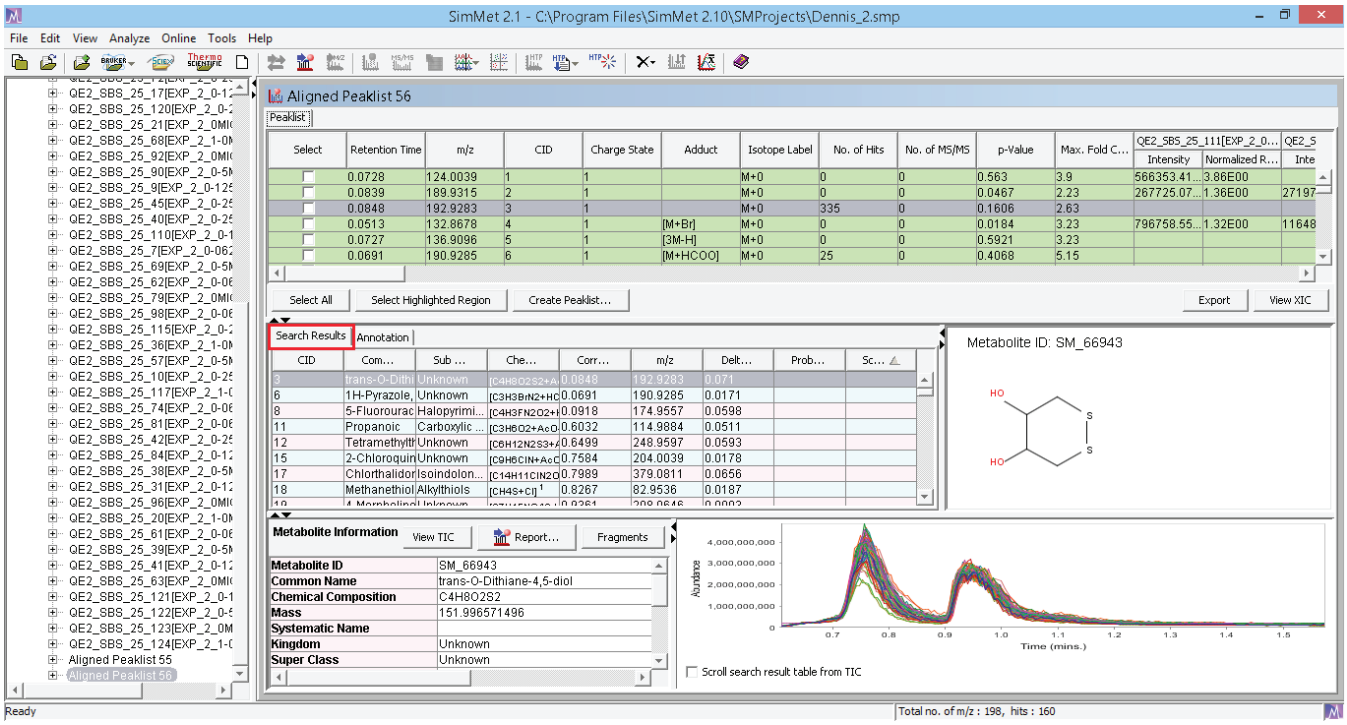


Figure 2: Typical graphical user interface of SimLipid software: A single workbench view of LC-MS peaklist, list of identified lipids at retention time points, structure of a lipid at a selected retention time point, and chromatogram of the sample with a vertical line indicating the retention time point of the displayed lipid structure

2. Metabolite Identification: Metabolite profiling using MS data is performed by exact mass database search. Further validation of metabolites can be done by matching observed MS/MS data against the SimMet database containing experimental MS/MS data of standard metabolites. A score determining the proximity between a candidate metabolite and a database metabolite is presented to users. The score is based on (a) the characteristic fragment ions generated by the candidate and the database metabolite, and (b) pattern matching between the observed MS/MS spectrum and the database MS/MS spectrum. SimMet database includes the NIST MS/MS library - 234,284 spectra: 51,216 ion trap spectra for 42,126 different ions of 8,171 compounds, 83,068 collision cell spectra (Q-TOF and tandem quad) and spectra for 14,835 different ions of 7,692 compounds.

SimMet®

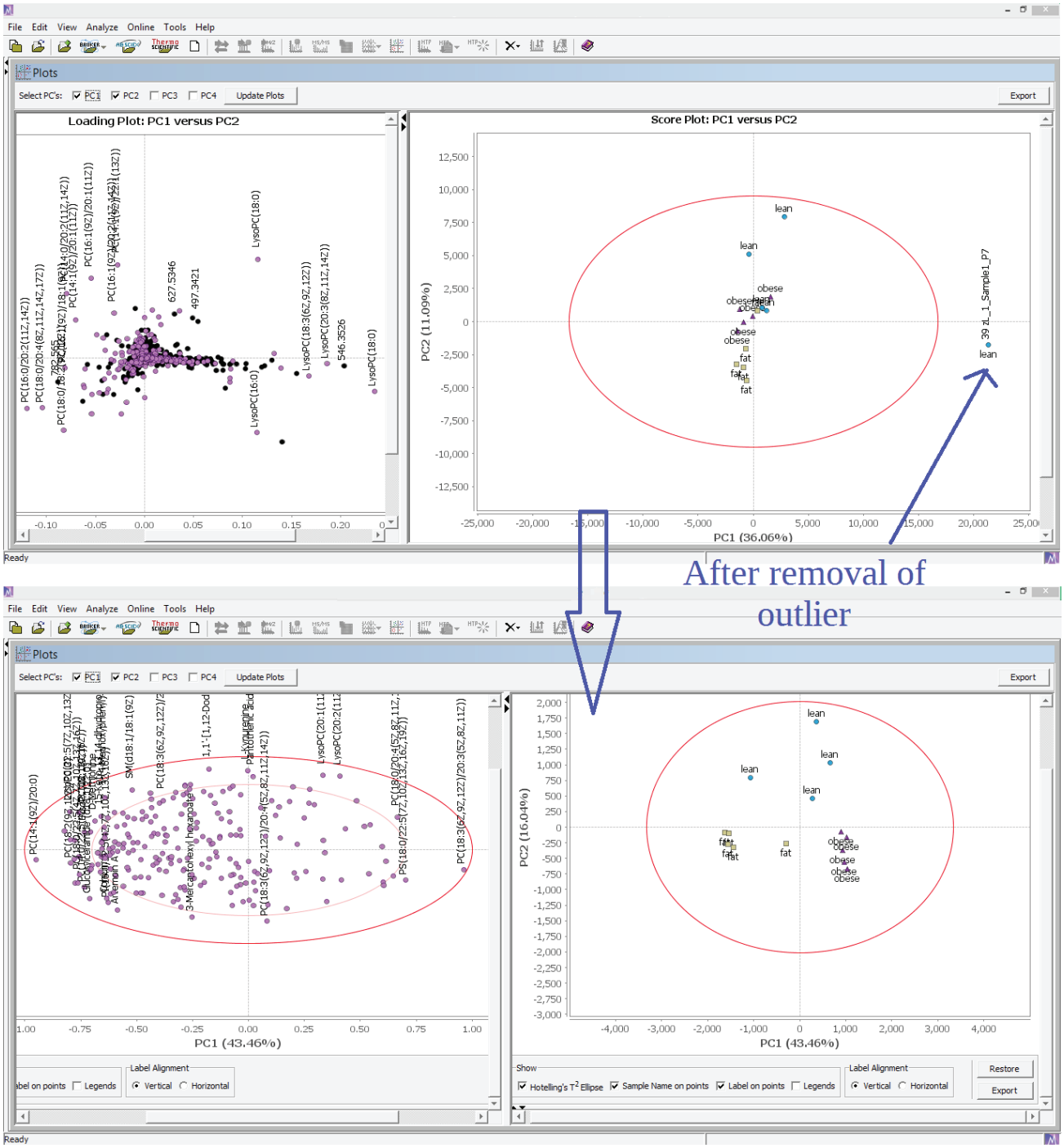
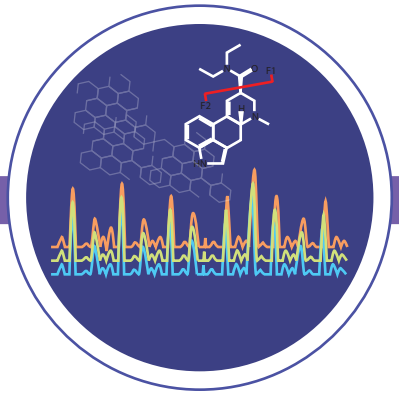


Figure 3: Typical graphical user interface of SimMet software: Principal Component Analysis (PCA) generates two 2-D plots, Scores plot and Loadings plot which help in identification of outliers. In order to pinpoint probable biomarkers, the 2-D plots are supported with Confidence Ellipses such as Correlation Loadings and Hotelling's T2 ellipses



3. Metabolite Differential Analysis: Differential metabolites across biological samples are identified using statistics such as fold change and p-value from ANOVA, and t-test. Furthermore, Principal Component Analysis can be performed to investigate variation and bring out strong patterns in a dataset. Loading and score plot coupled with confidence ellipses (correlation loadings and Hotelling's T2 ellipses) enable visualization of the relationship between metabolites and samples. Metabolite response curves are also provided to display the change in abundance of metabolites in different samples. High Throughput metabolite search after outlier removal further optimizes search results.

Portable Reports: Export data analysis results to customized reports in HTML, CSV, MS excel format files for sharing information with colleagues or publishing data.