About OMICS Group

OMICS Group International is an amalgamation of Open Access publications and worldwide international science conferences and events. Established in the year 2007 with the sole aim of making the information on Sciences and technology 'Open Access', OMICS Group publishes 500 online open access scholarly journals in all aspects of Science, Engineering, Management and Technology journals. OMICS Group has been instrumental in taking the knowledge on Science & technology to the doorsteps of ordinary men and women. Research Scholars, Students, Libraries, Educational Institutions, Research centers and the industry are main stakeholders that benefitted greatly from this knowledge dissemination. OMICS International also organizes 500 International conferences annually across the globe, where knowledge transfer takes place through debates, round table discussions, poster presentations, workshops, symposia and exhibitions.

About OMICS International Conferences

OMICS International is a pioneer and leading science event organizer, which publishes around 500 open access journals and conducts over 300 Medical, Clinical, Engineering, Life Sciences, Pharma scientific conferences all over the globe annually with the support of more than 1000 scientific associations and 30,000 editorial board members and 3.5 million followers to its credit.

OMICS International has organized 500 conferences, workshops and national symposiums across the major cities including San Francisco, Las Vegas, San Antonio, Omaha, Orlando, Raleigh, Santa Clara, Chicago, Philadelphia, Baltimore, United Kingdom, Valencia, Dubai, Beijing, Hyderabad, Bengaluru and Mumbai.

SimMet®

SimMet: Informatics Tool for Automating LC-MS and MS/MS Based Large Metabolomics Data Processing and Analysis

By

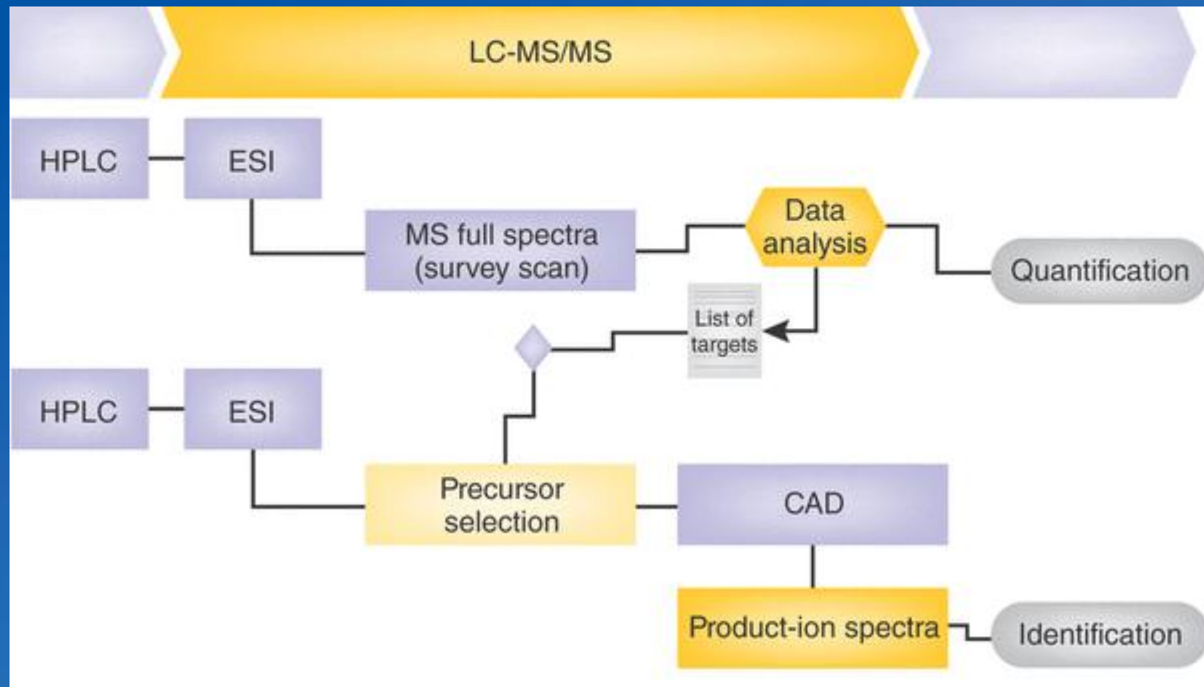
Sanjib N. Meitei, Ph.D.
Chief Scientific Officer, PREMIER Biosoft

at

4th International Conference and Exhibition on Metabolomics and System Biology, Philadelphia

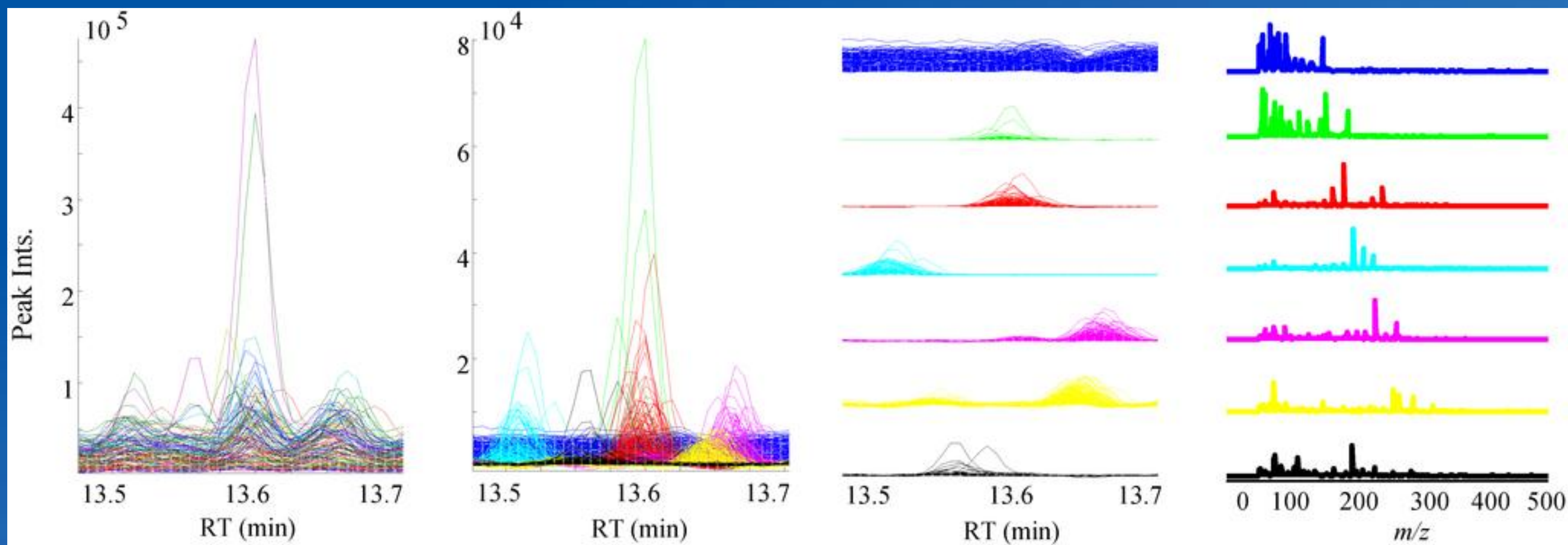
PREMIER Biosoft

Leading Analytical Platform Applied for Metabolite Profiling



Liquid Chromatography - Mass spectrometry (LC-MS)
Reason: High sensitivity and requirement of low sample amounts

Challenge: Large Data Sets



Typical Metabolomics Data Analysis Pipeline

1. Generate peak lists using a data processing tool.

2. Metabolite profiling using Database search tool.

3. Validate metabolites using MS/MS data pattern matching or *in silico* fragment matching tools.

4. Performing statistical analysis for identifying differential metabolites.

5. Pathways analysis for the identified metabolites.

Lack of a comprehensive software tool to perform all the steps mentioned above has been one of the bottle necks. In order to address this challenge, we develop SimMet software.

SimMet Schema

1. Import data from either native file formats and standard file formats

2. SimMet analyzes 500000 spectra from hundreds of biological samples in a single project

3. Data Processing
a. Peak Detection
b. Charge State Identification
c. Adduct Identification
d. Peak Alignment
e. Statistical Analysis to Pin Point Peaks of Interest

4. Metabolite Identification

a. MS1 Data Analysis
Remove Precursor Masses with No Match in Exact Mass Search

b. Isotope Pattern Matching
Remove Metabolites with IMS Less than a Threshold Value

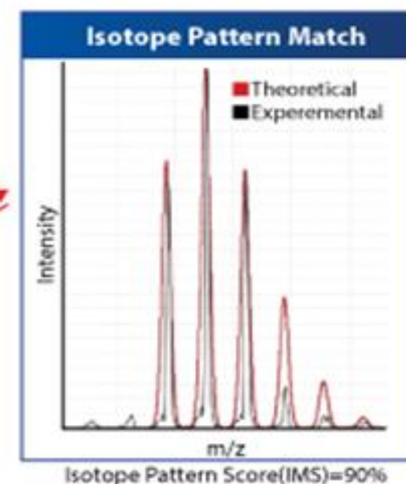
c. Product Ions or MS/MS Data Validation
Rank Candidate Metabolites of MS/MS Spectra Based on Probability Score

SimMet Database
MS1 Database

Exact Mass Search

Return IMS

Isotope Pattern Match



Isotope Pattern Score(IMS)=90%

SimMet Database
MS2 Database
NIST MS/MS and NIST MS/MS2 DBs (9344 cmpds 234284 MS/MS spectra and 45298 ions)

Match Observed Product Ions Vs Theoretical Diagnostic Ions

5. Identified Metabolites

6. Annotated Spectra
Label
a. MS Spectra with Matched Metabolite Structures
b. MS/MS Spectra with Matched Fragments

7. Quantify the Identified Metabolites

Export Labeled Spectra

8. Portable Reports
Export Results into MS Excel, HTML, CSV, JPEG and PNG File Formats

Export Results

Export Results

SimMet Software Addressing Some of The Challenges

S.No.	Challenges	SimMet
1.	Generation of a large number of LC-peaks from raw data	Molecular feature finding algorithm that combines peaks from adducts and higher isotopes
2.	Significant number of peaks corresponding to noise	Remove noise based on shape of the LC-peaks and data from LC-MS runs of blanks, QCs, technical replicates of the biological samples
3.	Missing peaks	Extracting raw data corresponding to a peak that are observed in other technical replicates
4.	Accurate identification of metabolites using MS/MS data	NIST MS/MS Database (having 234284 spectra, 9344 compounds and 45298 ions) and proprietary compound identification algorithm and spectral pattern matching algorithm

Application of Software (Methods)

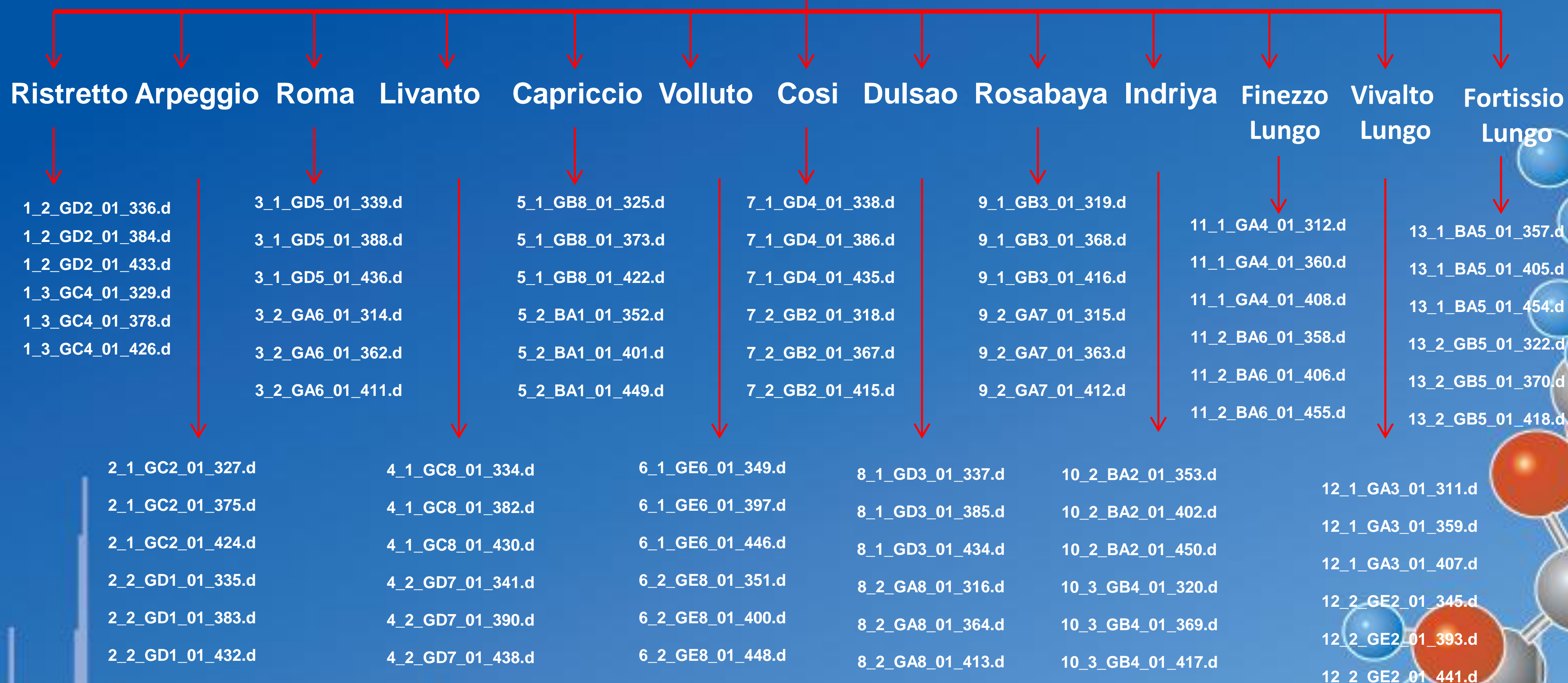
MS: Compact (Q-TOF MS, Bruker Daltonik GmbH). ESI(+) with MS and autoMS/MS modes. Scan range: m/z 75-1000. Acquisition rate: 3 Hz.

HPLC: U3000 RSLC(Thermo Scientific). Column: 50 x 2.1 mm BEH C18, 1.7 um column (Waters) Column temp. 30 °C. Flow rate: 0.45 mL/min. Injection volume: 5 µL. Mobile phase: A = H2O, B = MeOH (each containing 0.1% HCOOH). Gradient: linear gradient 2 - 98% B in 5 min, hold 1 min.

Sample: Capsules of 13 different types of coffee (espresso and lungo varieties from different blends and geographical regions) were extracted using 35 ml of water on a standard coffee capsule machine (Krups XN 301T Nespresso Pixie). Two replicates of each type were prepared. Extracts were diluted 1:50 in water prior to analyzing 3 replicates for each extract by UHPLC-MS.

Data Processing: SimMet software tool (www.premierbiosoft.com).

Coffee Samples



SimMet®

SimMet Data Analysis Workflow

Raw Data: Bruker's native files viz., .fid, .baf and .yep. profile data or line data types are supported. Other files: SCIEX's .wiff, Thermo's .raw, .zXML, .mzData.

SimMet 1.0 - C:\Program Files\SimMet\SMProjects\Coffee.smp

File Edit View Analyze Online Tools Help

Coffee
- RAW Data
- Database Search
- LC-MS and MS/MS Data
- Differential Analysis

(a)

Open Bruker Daltonics Peak

Look in: Coffee_13Samples

Recent Items, Desktop, Documents, This PC, Network

10_2_BA2_01_353.d, 10_2_BA2_01_402.d, 10_2_BA2_01_450.d, 10_3_GB4_01_320.d, 10_3_GB4_01_369.d, 10_3_GB4_01_417.d, 11_1_GA4_01_312.d, 11_1_GA4_01_360.d, 11_1_GA4_01_408.d, 11_2_BA6_01_358.d, 11_2_BA6_01_406.d, 11_2_BA6_01_455.d, 12_1_GA3_01_311.d, 12_1_GA3_01_359.d, 12_1_GA3_01_407.d, 12_2_GE2_01_345.d, 12_2_GE2_01_393.d, 12_2_GE2_01_441.d, 13_1_BA5_01_357.d, 13_1_BA5_01_405.d, 13_1_BA5_01_454.d, 13_2_GB5_01_322.d, 13_2_GB5_01_370.d, 13_2_GB5_01_418.d, 1_2_GD2_01_336.d, 1_2_GD2_01_384.d, 1_2_GD2_01_433.d, 1_3_GC4_01_329.d, 1_3_GC4_01_378.d, 1_3_GC4_01_426.d, 2_1_GC2_01_327.d, 2_1_GC2_01_375.d, 2_1_GC2_01_424.d, 2_2_GD1_01_335.d, 2_2_GD1_01_383.d, 2_2_GD1_01_432.d, 5_2_BA1_01_449.d

Folder name: :t\Sample data\Aiko\Coffee_13Samples\9_2_GA7_01_412.d"

Files of type: baf (baf)

Open | Cancel

Load Scan

(b)

Data Type
- Profile data
- Line data
- Import peak tops
- If selected data type is not available for any scan, import available profile/line data
- Select precursor m/z from parent MS data — m/z variation: 0.5 Dalton

Fill all blank fields

Precursor Ion m/z: | Charge State: | Apply

4_2_GD7_01_341.d

Show MS Only | Show MS with MS/MS Only | Show All

Spectrum No.	Retention Time (secs.)	MS Level	Precursor m/z	Charge State	Polarity	Select Scan
1	0.928	1		1	Positive	
2	1.268	1		1	Positive	
3	1.608	1		1	Positive	
4	1.948	1		1	Positive	
5	2.288	1		1	Positive	
6	2.628	1		1	Positive	

Reset | Specify Range | Select All | Clear All

Select All Scans of All Profiles | Specify Range | Load Scan | Cancel | Help

Opens Bruker Daltonics baf/yep files

SimMet®

Model Experimental Design Through Intuitive Interfaces

A software-wizard that guides users to model experimental design by assigning raw data files to their respective biological/technical replicates, assign color code, shape and custom description for each of the biological/technical replicates.

Model Experimental Design

Select: Raw Files / Peaklists

File Name
1_3_GC4_01_426.d_P90
2_1_GC2_01_327.d_P91
2_1_GC2_01_375.d_P92
2_1_GC2_01_424.d_P93
2_1_GC2_01_424.d_P93
2_1_GC2_01_424.d_P93
2_1_GC2_01_424.d_P93
3_1_GC2_01_424.d_P93
3_2_GD1_01_335.d_P94

Sample Name: Arpeggio

Amount of Sample (µg/mol): 1.0

Polarity: Positive

>> <<

Color 1

Colors

Edit colors

File Name	Sample Name	Description	Amount	Polarity	Color...	Shape
2_1_GC2_01_327.d_...	Arpeggio	Replicate 1	1.0	Positive		
2_1_GC2_01_375.d_...	Arpeggio	Replicate 2	1.0	Positive		
2_1_GC2_01_424.d_...	Arpeggio	Replicate 3	1.0	Positive		
2_1_GC2_01_424.d_...	Arpeggio	Replicate 4	1.0	Positive		
2_1_GC2_01_424.d_...	Arpeggio	Replicate 5	1.0	Positive		
2_1_GC2_01_424.d_...	Arpeggio	Replicate 6	1.0	Positive		
3_1_GC2_01_424.d_...	Roma	Replicate 1	1.0	Positive		

Data Normalization

Lock Parameters

Peak Detection & Picking
Template Name: Compact Qq-TOF
Settings...

Align Peaklists
Template Name: Compact Qq-TOF
Settings...
Aligned Peaklist Name: Coffee_Compare

Metabolite Identification
Template Name: Compact Qq-TOF
Settings...

Statistical Analysis
PCA
PLS-DA
Hierarchical Cluster Analysis

OK Cancel Help

PREMIER Biosoft

SimMet®

Model Experimental Design Through Intuitive Assistance Dialogs

Define Data Analysis Pipeline: Peak detection and picking, feature detection, retention time alignment, metabolite identification using MS and MS/MS database search, and statistical analysis such as Principal Component Analysis (PCA), Partial Least Square Discriminant Analysis (PLS-DA), Hierarchical Cluster Analysis etc.

Data Normalization: Select proper data pretreatment method.

Data Normalization, Centering, Scaling & Transformation

Select Technique:

Response ○ Area ◉ Intensity

☑ Select a data normalization technique
- ◉ None
- ○ Normalization by internal standards
- ○ Normalization by total response sum
- ○ Normalization by median
- ○ Normalization by manual sample wise factors

☑ Normalization by a reference Sample [Roma]

☑ Data centering/scaling
- ○ None
- ○ Mean Centering
- ◉ Autoscaling (mean-centered and divided by the standard deviation of each variable)
- ○ Pareto Scaling (mean-centered and divided by the square root of standard deviation of each variable)
- ○ Range Scaling (mean-centered and divided by the range of each variable)
- ○ VAST Scaling (Variable Stability: Autoscaling multiplied by inverse of coefficient of variation)
- ○ Level Scaling (mean-centered and divided by mean of each variable)

☑ Data transformation

○ None ◉ Log transformation ○ Power transformation

OK Cancel Help

Raw Data Processing Steps

Generate Peaklists in Batch: Peaks detected in LC timescales for hundreds of raw data files in batch mode.

Peak Deconvolution: Separate isomeric/isobaric compounds by subjecting Extracted ion chromatogram (XIC) data into second derivative Savitzky-Golay smoothing.

Data Reduction: Combines all ions belonging to the same compound (peaks corresponding to isotopes, charge states, adducts and common neutral losses such as, NH4, Na, Li, K etc.)

Compound ID: A unique ID for the detected compound. All MS/MS scans corresponding to ions of this ID are also clustered.

Retention Time Alignment: Either RANSAC or Gale-Shapely techniques.

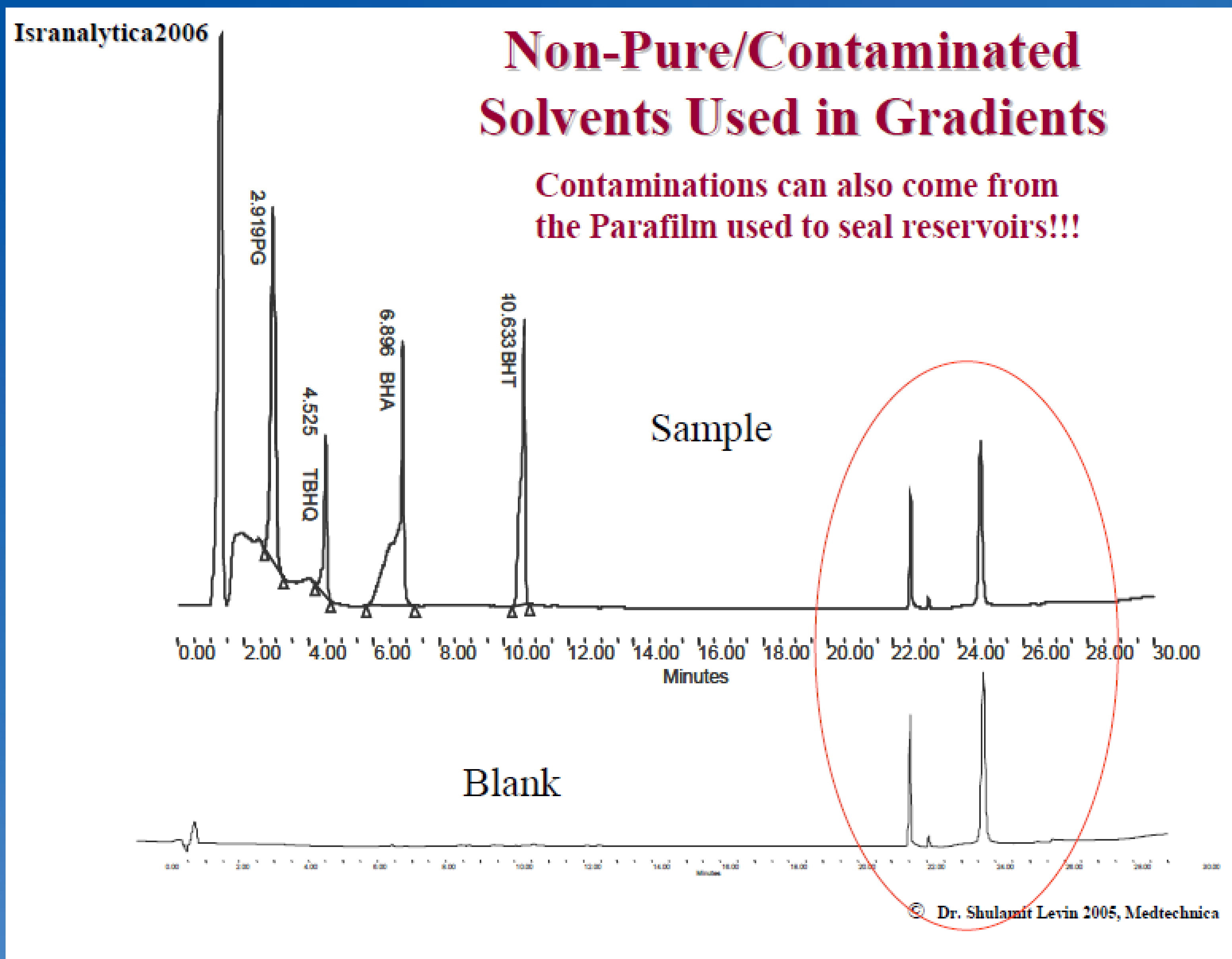
Review Peaks: Options to remove unwanted peaks, fetch intensity from raw data files for missing peaks.

Removing Noise Using Blank Samples

1. LC-MS run of blank extracts subjected to peak detection and picking and then aligned based on retention time with other sample peaklists.

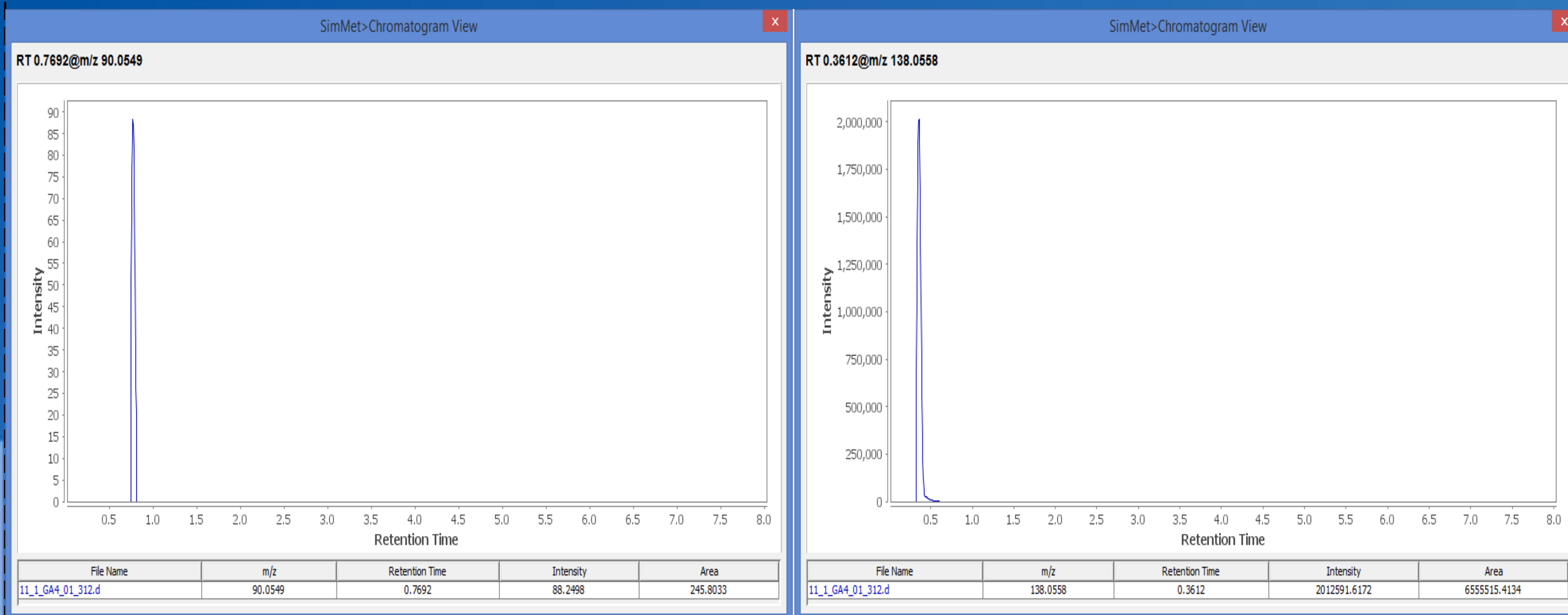
2. All the peaks that are aligned with peaks detected in the blank extracts are removed from further analysis.

Hence, unwanted peaks are removed without increasing the risk of removing compounds that have low abundances with poor signal to noise ratios.



Ability to Detect Metabolites with Very Low Concentration

Figures below show the XICs of the low concentrated alanine that was detected with an intensity of 88 cts versus the trigonelline peak that has an intensity of 2012591 cts. The ratio 2012591/ 88 = 2.2 X 10^4 > 4 orders of magnitude.

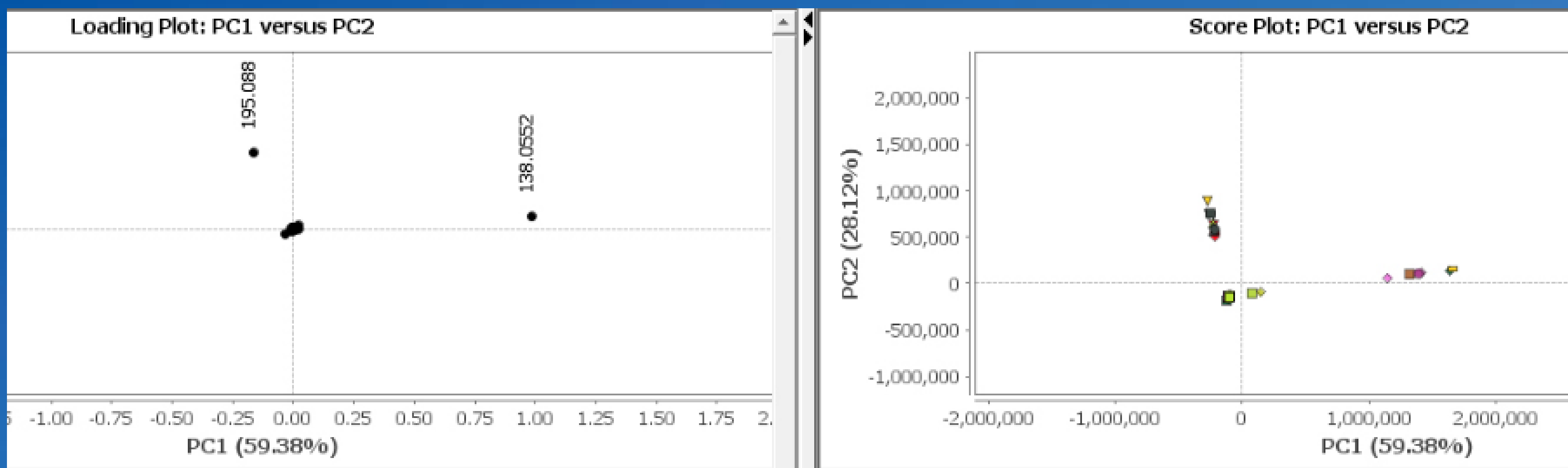


This observation also demonstrates the unique capability of the compact QTOF to detect target compounds on an LC timescale across a dynamic range.

Metabolite Differential Analysis

Principal Component Analysis: Aligned peaklists from13 different coffee extracts

PCA Score Plot: The 2 “biological” and 3 technical replicates for each sample type (highlighted by using the same color and symbol) formed clusters in the PCA scores plot as shown in figure below.



PCA Loadings Plot: Showing analytes with m/z values 195.088 and 138.0552 mainly contributing to the separation of samples in the PCA scores plot . As m/z 195.088 corresponds to caffeine, we removed it from the model and re-run the PCA data analysis. Two compounds labeled X with m/z value 124.0394 and Y with m/z value 138.0561 are detected to have a high content in strong and weak coffee samples, respectively .

Metabolite Identification

Basics of Compound Identification: Exact mass database search, and matching of expected and observed fragments in batch mode.

Identify Candidate Metabolites: Search metabolites with precursor m/z as the search predicate and use 5ppm as the tolerance.

Identify the Most Likely Structure of the Candidate: Standard MS/MS spectra for all the candidates are matched to the measured spectra with user specified tolerance. A proprietary ranking algorithm differentiates isobaric compounds based on the number of matched observed ions and intensities of those matched ions.

Scoring Mechanism: A propriety algorithm that assigns penalty for fragment ions that can not be matched to database ions wherein the amount of penalty is decided based on the relative intensity of the non interpreted ions. The higher a penalty a structure receives, the lower the likelihood that the structure corresponds to the MS/MS spectrum.

Portable Reports: MS excel, CSV and HTML files.

SimMet®

Identification of Caffeine using MS and MS/MS Data

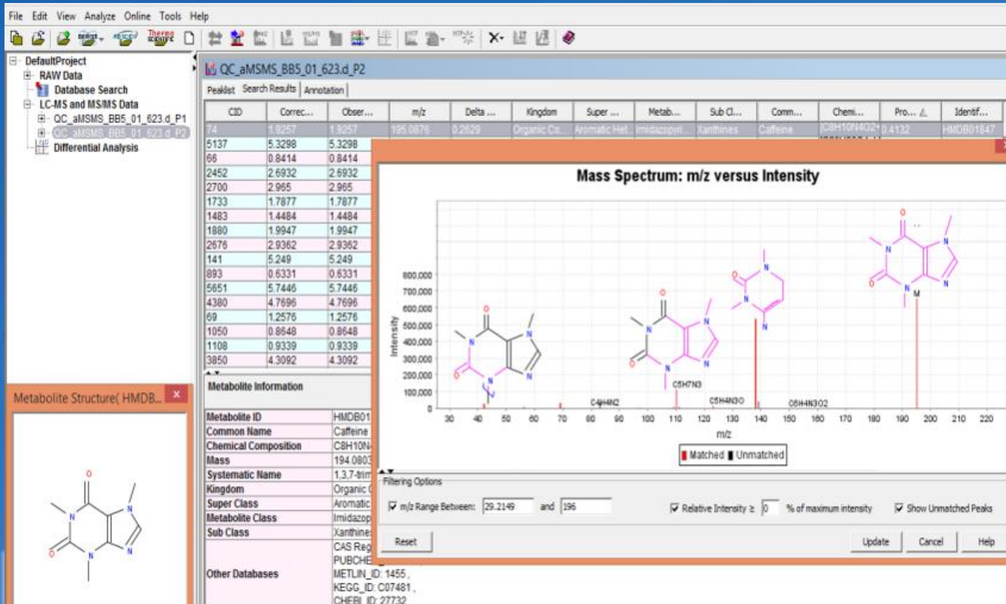
Goal: Test SimMet software's ability to accurately identify metabolite using MS and MS/MS data

Caffeine MS/MS Data: The QC sample data subjected into SimMet's MS and MS/MS database search workflow.

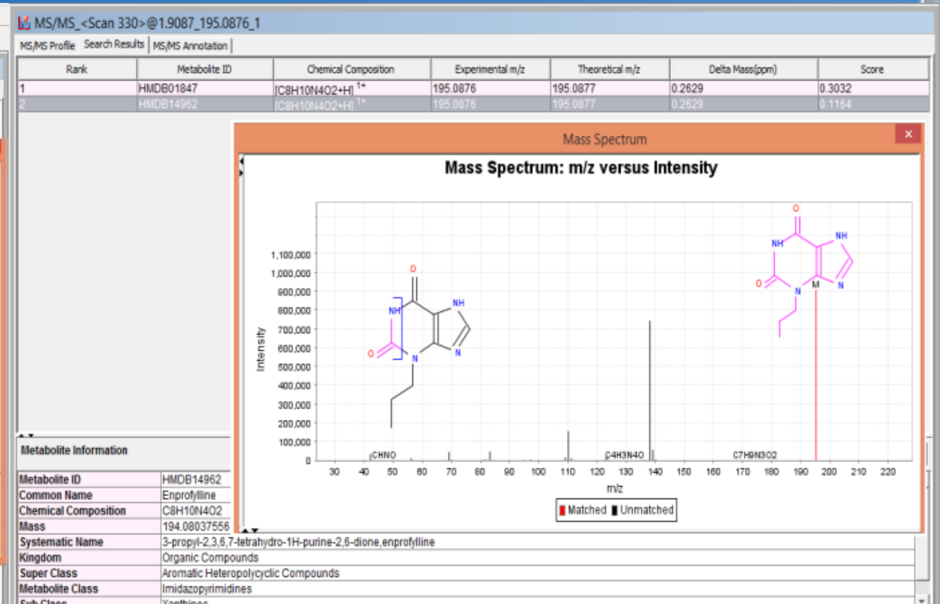
Compounds Identified: Caffeine structure was correctly identified and ranked 1st. Enprofylline was the 2nd ranked structure.

MS/MS Annotation Showing Fragmentation Patterns of IDed Compounds: Caffeine (Figure (a)) enprofylline in Figure (b). It is evident that many important diagnostic ions such as m/z 138, 110, 42 etc. are left unmatched in (b).

(a)



(b)



Identify Compounds X and Y Observed in PCA Score Plot

Candidate Structure: Exact mass search with 5 ppm tolerance in metabolite databases such as HMDB, YMDB, Metlin.

Compounds ID using Minimum Delta Mass Only: Compounds X tentatively identified as Isonicotinic acid and compound Y as 4-Fluoro-L-threonine

MS/MS Data: We use MS/MS data from QC sample.

Averaging MS/MS scans for the Compound X: The MS/MS scans observed between 0.4-0.6 minute, the LC-timerange in which compound X is eluting.

Average Scans

Average Scans

MS Data   MS/MS Data

Profile Name: QC_aMSMS_BB5_01_623.d

Scan Name	Scan No.	Precursor m/z	Charge State	Intensity	Retenti...	
QC_aMSMS_BB5_0...	74	124.0389	1	7500.0	25.884	☐
QC_aMSMS_BB5_0...	72	124.039	1	4706.0	25.175	☑
QC_aMSMS_BB5_0...	90	124.0392	1	69797.0	31.424	☑
QC_aMSMS_BB5_0...	86	124.0392	1	102263.0	30.044	☑

Profiles/Scans Checked: 0

Filtering Options

Scan No: to   Precursor m/z: 124.03 to 124.05

Retention Time: 24 to 36 (secs.)

Ignore the row for which Retention Time/Precursor m/z is not available   Filter   Default

Data Pre-processing Options

m/z Options

Consider m/z peaks observed in all the samples   m/z Precision (decimal points): 4

Intensity Options

Multiple peaks with same m/z within roundoff limits: Threshold Intensity:

Average Intensity   Highest Intensity   Relative   Value 0.0 %

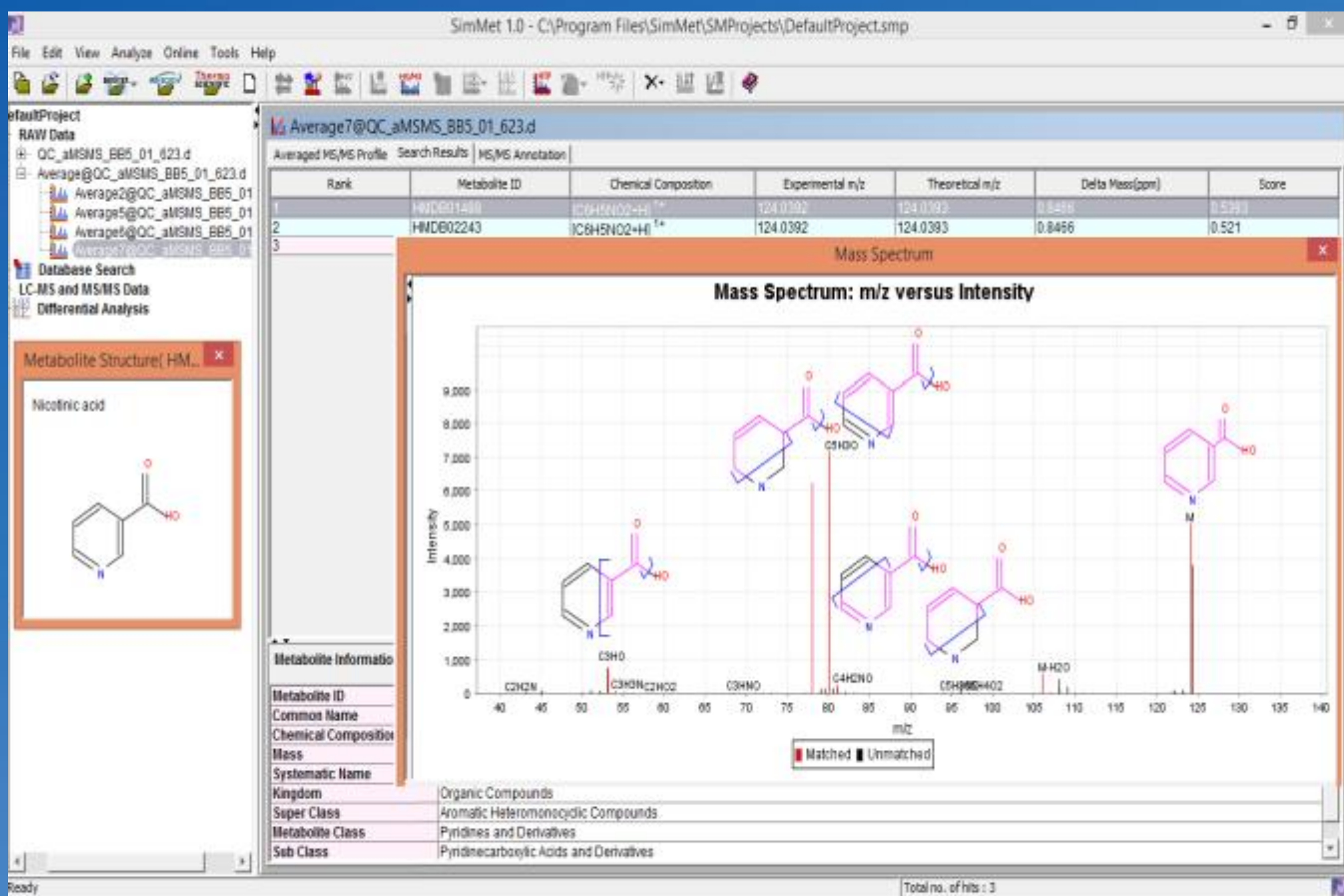
Average Scans   Cancel   Help

Identify Compounds X and Y Observed in PCA Score Plot

MS/MS based Identification of Compound X:
1st Ranked Structures: Nicotinic acid, 2nd Ranked: Picolinic acid and 3rd Ranked: Nitrobenzene.

MS/MS spectra annotation of compound X: As shown in figure below.

Using the workflow described here, the molecular formula for compound Y was identified
as C7H8NO2 ([M+H]+).
In-silico fragmentation identified the analyte to be trigonelline. The ID of compound X as nicotinic acid and compound Y as trigonelline was confirmed by comparison to the authentic standards.



Conclusion

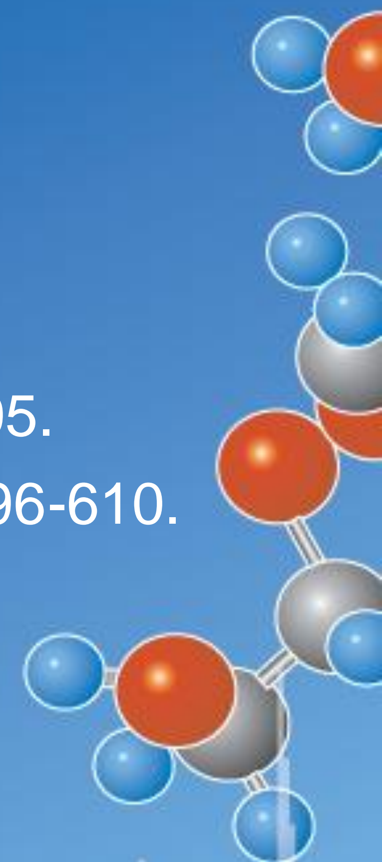
SimMet, a high throughput sophisticated software for comprehensive LC-MS and MS/MS metabolomics data analysis, enables accurate detection of peaks, quick pinpointing of relevant compounds contributing to coffee intensity, identification of two selected target compounds which are characteristic for weak and strong coffee samples.

The complete data analysis of the data set could be performed on a single software platform.

This reliable proposal of compound identities helped to save analysis time and money spent for purchasing multiple references in order to confirm the identity of the target compounds.

References

1. Castrillo et al. Phytochemistry. 2003; 62: 929–37.
2. Theodoridis et al. Mass Spectrom Rev. 2011; 30: 884–906.
3. Bajad et al. Methods Mol Biol. 2011; 708: 213–28.
4. Xia et al. Nucl. Acids Res. 2012; 37: W652-660.
5. Gowda et al. 2014; 86 (14): 6931-6939.
6. Eric et al. Nat Rev Genet. 2010; 11(9): 647–657.
7. Berg et al. BMC Genomics. 2006; 7:142.
8. Fischler and Bolles. Comm Of the ACM. 1981; 24:381-395.
9. Cleveland and Devlin. J Am Stat Assoc. 1998; 83(403):596-610.
10. Pluskal et al. BMC Bioinformatics. 2010; 11:395.
11. Voss et al. Bioinformatics. 2011; 27(7):987-93.
12. Bruker Application Note # LCMS-79



Acknowledgement

i.A. Dr. Aiko Barsch
Market Manager Metabolomics
Bruker Daltonik GmbH (Bremen)
Email: aiko.barsch@bruker.com

And
Development Team, PREMIER Biosoft

Let Us Meet Again

We welcome you all to our future conferences of

OMICS International

Please Visit:

www.metabolomicsconference.com

www.conferenceseries.com

http://www.conferenceseries.com/clinical-research-conferences.php