LC-MS and MS/MS based Large Metabolomic Data Processing and Analysis using SimMet®

Ningombam Sanjib Meitei[1], Anith Kumar Sreenidhi[1], Arun Apte[2], Baljit K Ubhi[3]
[1]PREMIER Biosoft, Indore, India, [2]PREMIER Biosoft, Palo Alto, U.S.A, [3]SCIEX, Redwood City, CA, USA.
Corresponding author e-mail: sanjib@premierbiosoft.com

PREMIER Biosoft

1

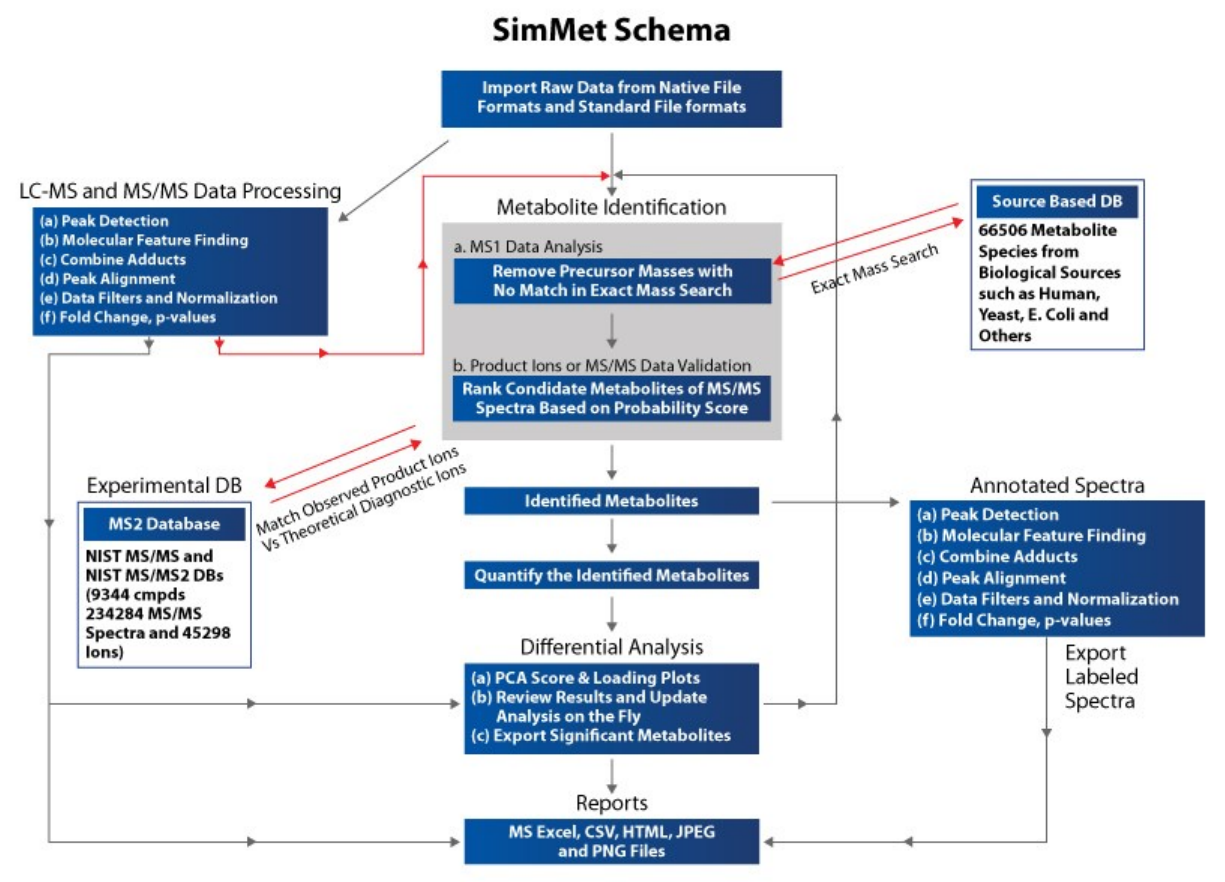
Introduction

Despite rapid growth in innovations related to liquid chromatography-tandem mass spectrometry (LC-MS/MS) based metabolite profiling studies [1-3], lack of high throughput software tools has been one of the bottlenecks. A typical metabolomics data analysis pipeline may include multiple software tools for example, using of (1) a data processing tool to generate peaklists (ii) a database search tool for metabolite profiling, (iii) other tools to validate metabolites using MS/MS data pattern matching or *in silico* fragment matching, (iv) tools for performing statistical analysis for identifying differential metabolites and (v) pathway analysis for the identified metabolites. In order to address the challenges, we have developed SimMet. The software workflow will be demonstrated based on a Zucker rat study well characterized for studying effects of obesity, diabetes and cardiovascular effects.

Methods

Using a HPLC system and a high strength silica column (Acquity HSS T3 1.8µm, 2.1 x 100mm @ 60°C), polar metabolites from a 5uL injection of serum were separated at a flow rate of 600 µL/min. Full scan TOF MS and MS/MS data was acquired simultaneously on a TripleTOF® 5600+ system in DDA, single injection workflow (SCIEX). Serum from the Zucker rat model was taken from 7-9 week old lean, fatty and obese rats (n= 10/sample). Samples were methanol extracted to remove the protein, centrifuged and the dried extract diluted in 100µl of Water: Methanol (80:20). A pooled sample was used as a QC (acquired every 5 samples) to monitor data reproducibility.

SimMet software has been used for data analysis. Figure 1 shows the schematic representation of the software work flow. Software protocols consist of major features namely data processing, metabolite identification and differential analysis. After importing raw LC-MS and MS/MS data, peaks were detected in LC timescales. The program employs a feature finding algorithm (FFA) to identify adducts, isotopes and charge states of a detected compound and combines all the corresponding LC-peaks under a unique Compound ID (CID) number. Figure 2 shows a typical MS spectrum of a detected compound with peaks annotated by corresponding adducts.



SimMet Schema

Import Raw Data from Native File Formats and Standard File formats

LC-MS and MS/MS Data Processing
- (a) Peak Detection
- (b) Molecular Feature Finding
- (c) Combine Adducts
- (d) Peak Alignment
- (e) Data Filters and Normalization
- (f) Fold Change, p-values

Metabolite Identification
- a. MS1 Data Analysis: Remove Precursor Masses with No Match in Exact Mass Search
- b. Product Ions or MS/MS Data Validation: Rank Condidate Metabolites of MS/MS Spectra Based on Probability Score

Exact Mass Search

Source Based DB: 66506 Metabolite Species from Biological Sources such as Human, Yeast, E. Coli and Others

Experimental DB: 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

Identified Metabolites

Quantify the Identified Metabolites

Differential Analysis
- (a) PCA Score & Loading Plots
- (b) Review Results and Update Analysis on the Fly
- (c) Export Significant Metabolites

Annotated Spectra
- (a) Peak Detection
- (b) Molecular Feature Finding
- (c) Combine Adducts
- (d) Peak Alignment
- (e) Data Filters and Normalization
- (f) Fold Change, p-values

Export Labeled Spectra

Reports: MS Excel, CSV, HTML, JPEG and PNG Files

Figure 1: Schematic representation of SimMet software.

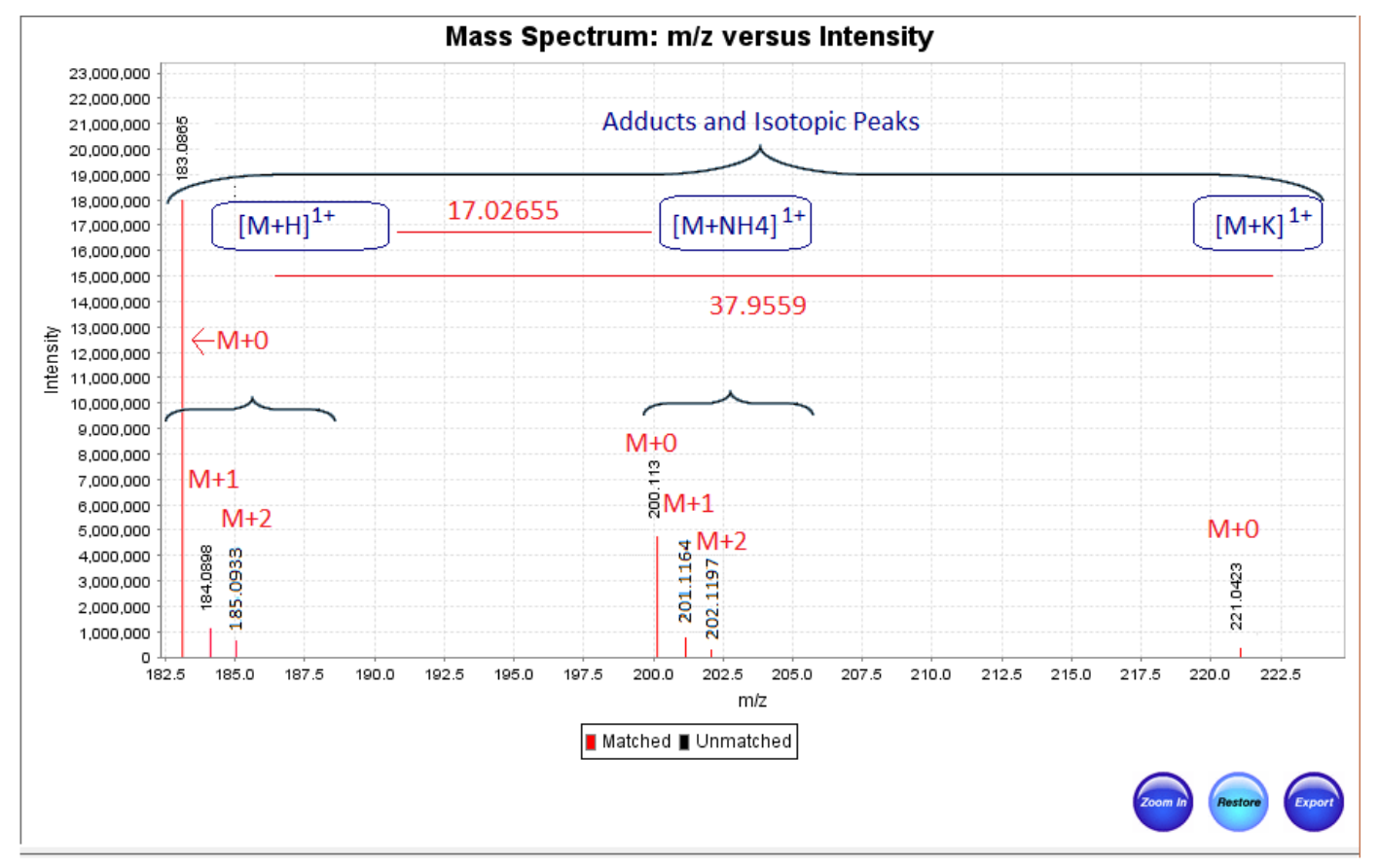


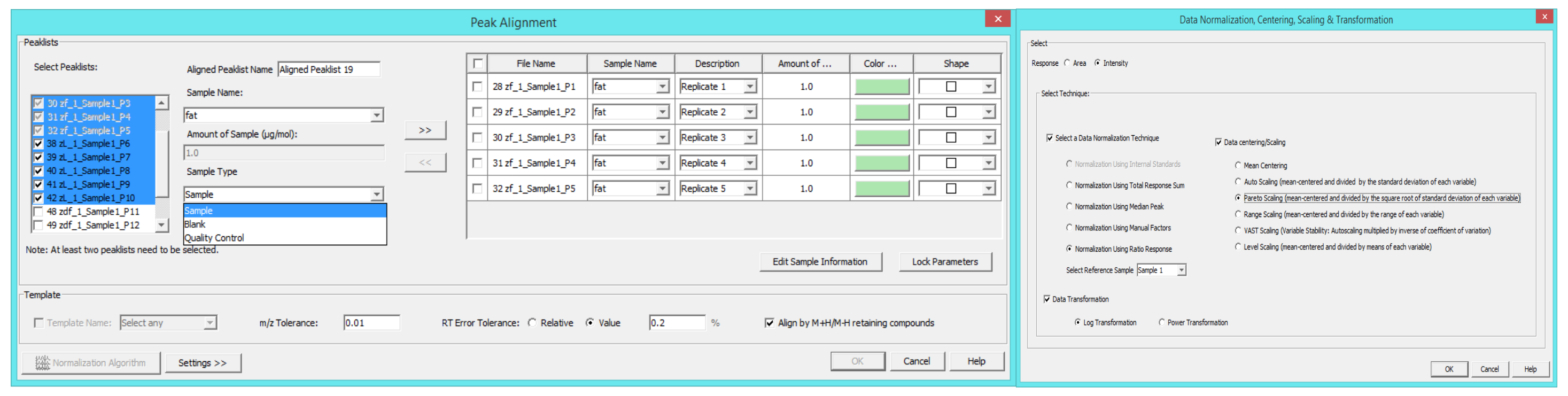
Figure 2: Typical SimMet output dialog showing annotated LC-MS peaks of a detected compound.

Further, for each CID, the program identifies all the MS/MS scans that have been acquired between the start and end time. If precursor m/z value of an MS/MS scan corresponds to component peak with higher isotope level i.e., M+1, M+2, etc., then it is automatically replaced with the m/z value corresponding to its monoisotopic peak.

Model Experimental Design

Data can be grouped according to biological and technical replicates (Figure 3(a)). Peaklists can be categorized as blank, QC and samples. LC peaks from samples may be removed based on (non-) observation of the corresponding LC peaks in blanks and QCs. Appropriate data normalization and transformation methods can be specified (Figure 3(b)).

2



Peak Alignment

Select Peaklists: ☑ 30 zf_1_Sample1_P3, ☑ 31 zf_1_Sample1_P4, ☑ 32 zf_1_Sample1_P5, ☑ 38 zL_1_Sample1_P6, ☑ 39 zL_1_Sample1_P7, ☑ 40 zL_1_Sample1_P8, ☑ 41 zL_1_Sample1_P9, ☑ 42 zL_1_Sample1_P10, ☐ 48 zdf_1_Sample1_P11, ☐ 49 zdf_1_Sample1_P12

Aligned Peaklist Name: Aligned Peaklist 19
Sample Name: fat
Amount of Sample (µg/mol): 1.0
Sample Type: Sample (Sample / Blank / Quality Control)

File Name	Sample Name	Description	Amount of ...	Color ...	Shape
28 zf_1_Sample1_P1	fat	Replicate 1	1.0		□
29 zf_1_Sample1_P2	fat	Replicate 2	1.0		□
30 zf_1_Sample1_P3	fat	Replicate 3	1.0		□
31 zf_1_Sample1_P4	fat	Replicate 4	1.0		□
32 zf_1_Sample1_P5	fat	Replicate 5	1.0		□

Note: At least two peaklists need to be selected.

Template: ☐ Template Name: Select any; m/z Tolerance: 0.01; RT Error Tolerance: ○ Relative ● Value 0.2 %; ☑ Align by M+H/M-H retaining compounds

Data Normalization, Centering, Scaling & Transformation

Response ○ Area ● Intensity

☑ Select a Data Normalization Technique: ○ Normalization Using Internal Standards; ○ Normalization Using Total Response Sum; ○ Normalization Using Median Peak; ○ Normalization Using Manual Factors; ● Normalization Using Ratio Response; Select Reference Sample: Sample 1

☑ Data centering/Scaling: ○ Mean Centering; ○ Auto Scaling (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 means of each variable)

☑ Data Transformation: ● Log Transformation ○ Power Transformation

Figure 3: Typical SimMet software interfaces to (a) model experimental design (b) Select data normalization and transformation technique.

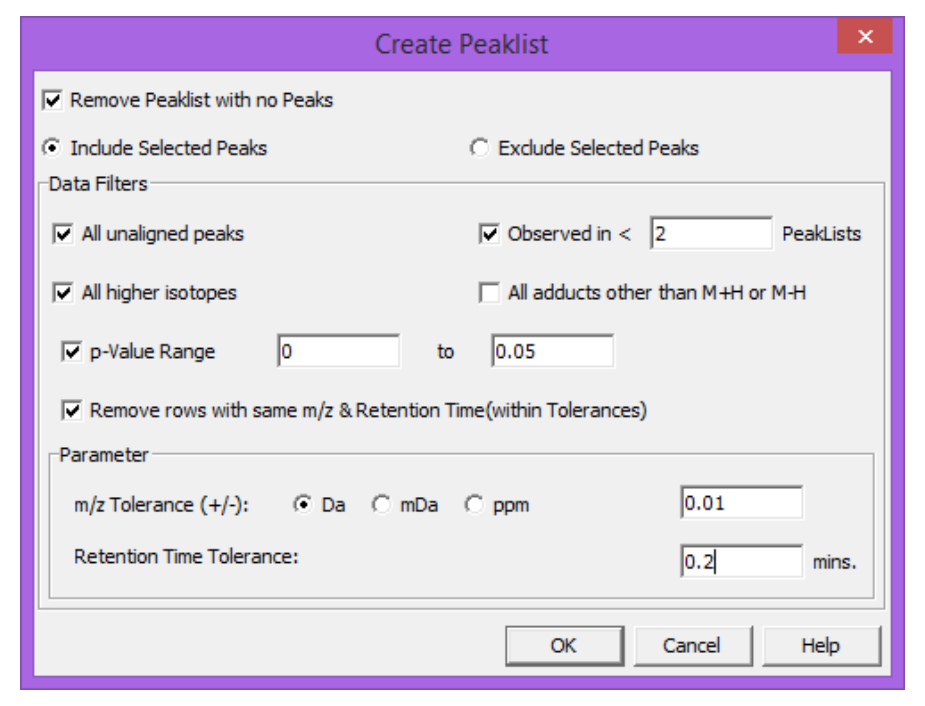
Retention Time Alignment

SimMet allows users to perform retention time alignment using one of the algorithms namely Stable Marriage[4] or RANSAC algorithm[5] in combination with locally-weighted scatterplot smoothing (LOESS)[6,7] method for regression which effectively align peaks with non-linear deviation of the retention times among samples. The accuracy and efficiency of this peak alignment model has been reported[8-9]. Missing peaks in the aligned data can be filled by extracting intensity of the peaks from raw data using the LC time scale of the detected peaks.

Statistical Analysis

For each detected compound, maximum fold change and p-value are reported. The p-value is calculated using either ANOVA test (or t-test) with the null hypothesis that abundance of metabolites across biological replicates do not vary. The calculated p-values and fold changes can be used as data filters (Figure 4) to create sub sets of the data.

The retention time-aligned data can be subjected to either MS or MS/MS database search for metabolite identification or differential analysis using Principal Component Analysis (PCA). PCA will enable unsupervised clustering of samples and pin-point the metabolites which significantly contribute to distinguish samples. MS or MS/MS database search for metabolite identification can be performed after PCA. Further, data normalization technique, experimental design, LC-peaks and peaklists can be changed on the fly to update the analysis. This flexibility enables removal of data outliers from the analysis, thereby leading to enhanced information.



Create Peaklist

☑ Remove Peaklist with no Peaks
● Include Selected Peaks ○ Exclude Selected Peaks

Data Filters
- ☑ All unaligned peaks
- ☑ Observed in < 2 PeakLists
- ☑ All higher isotopes
- ☐ All adducts other than M+H or M-H
- ☑ p-Value Range 0 to 0.05
- ☑ Remove rows with same m/z & Retention Time(within Tolerances)

Parameter
- m/z Tolerance (+/-): ● Da ○ mDa ○ ppm 0.01
- Retention Time Tolerance: 0.2 mins.

Figure 4: Typical SimMet software data filters.

Metabolite Identification

All the monoisotopic masses or corrected precursor m/z values (in case of MS/MS data) are subjected into exact mass database search (Database statistics provided in Figure 1) using program identified or user supplied ion species information. For each MS/MS data set, candidate metabolites with precursor m/z as the search predicate is listed. Standard MS/MS spectra for all the candidates (NIST MS/MS and MS/MS2) are matched to the observed 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. The algorithm assigns penalty for ions that cannot 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.

3

Results and Discussion

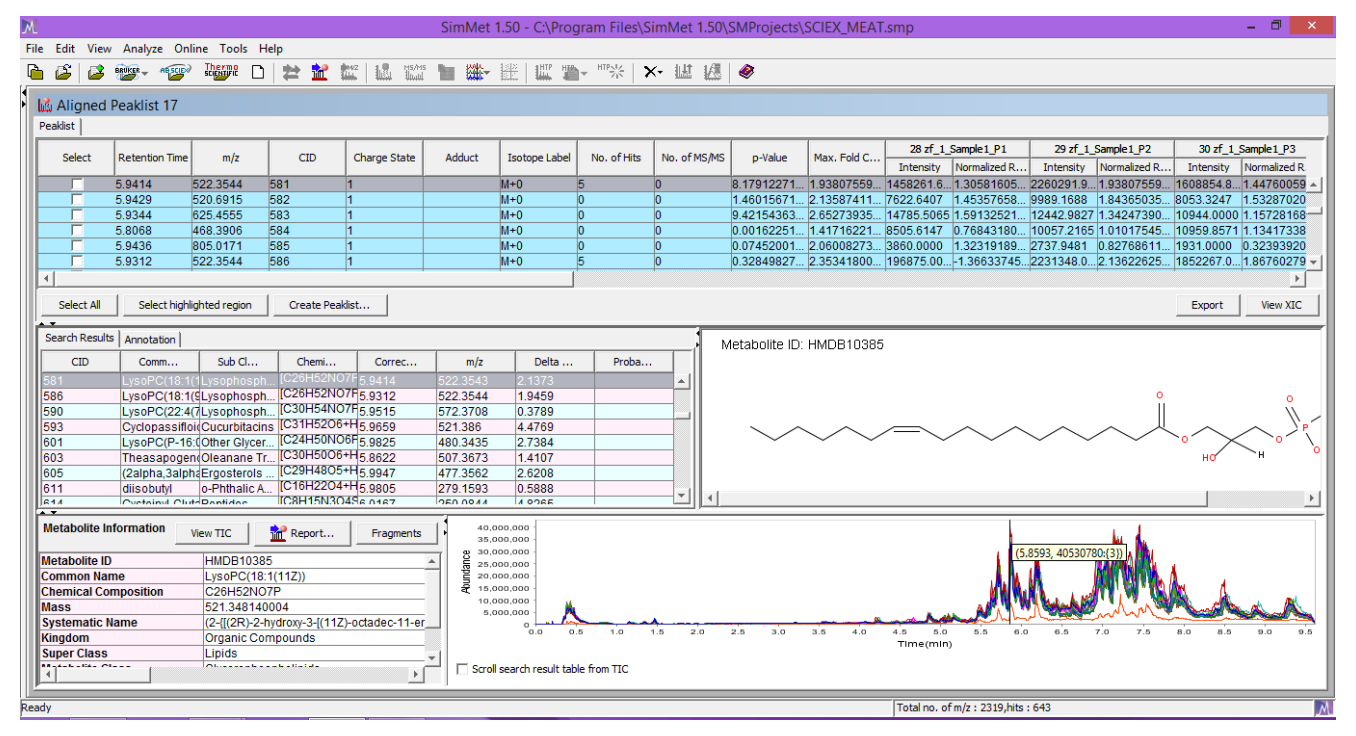
Application of SimMet to metabolite profiling and differential analysis of Zucker rat samples

On subjecting the raw files into SimMet's molecular feature finding and retention time alignment modules, 2319 compounds were identified out of which 411 compounds can be validated using MS/MS data (688 MS/MS scans). The retention time aligned data, identified metabolites, statistics, structures, metabolite information such as common name, class, etc., overlays of extracted ion chromatograms, total ion chromatogram, annotated MS or MS/MS spectrum for a metabolite can be visualized at a single workspace.

# cmpds by FFA module	# cmpds with MS/MS data	# MS/MS scans	# metabolites IDed	# metabolites with p-value <= 0.05
2319	411	688	643	228
			63 Metabolite classes	40 metabolite classes

Table 1: Summary of the LC-MS and MS/MS metabolic profile of the zucker rat samples. Detailed information on different classes and frequency of metabolites are provided in table 2.

User can scroll through the total ion chromatogram to locate the identified metabolites at a LC time point. For example, on selecting the time point 5.9 minutes in the TIC plot (Figure 5 (a), lower right plot), LysoPC 18:1 is automatically selected and each table in the figure 5(a) displays information regarding the identified lipid. The comparison plot between observed vs standard spectra of the metabolite from NIST MS/MS database is displayed on pushing the button 'Plot Spectrum' in the 'Annotation' tab (Figure 5(b)). XIC can be viewed by clicking View XIC button (Figure 6). Figure 7 shows the pie charts of percentage of identified metabolites species from different classes.



SimMet 1.50 - C:\Program Files\SimMet 1.50\SMProjects\SCIEX_MEAT.smp

Aligned Peaklist 17

Metabolite ID: HMDB10385

Metabolite Information	
Metabolite ID	HMDB10385
Common Name	LysoPC(18:1(11Z))
Chemical Composition	C26H52NO7P
Mass	521.348140004
Kingdom	Organic Compounds
Super Class	Lipids

Figure 5(a): Typical SimMet software interface displaying metabolic profile in a single workspace. Scroll through the TIC plot to locate the metabolite at a LC time point (lower right plot)

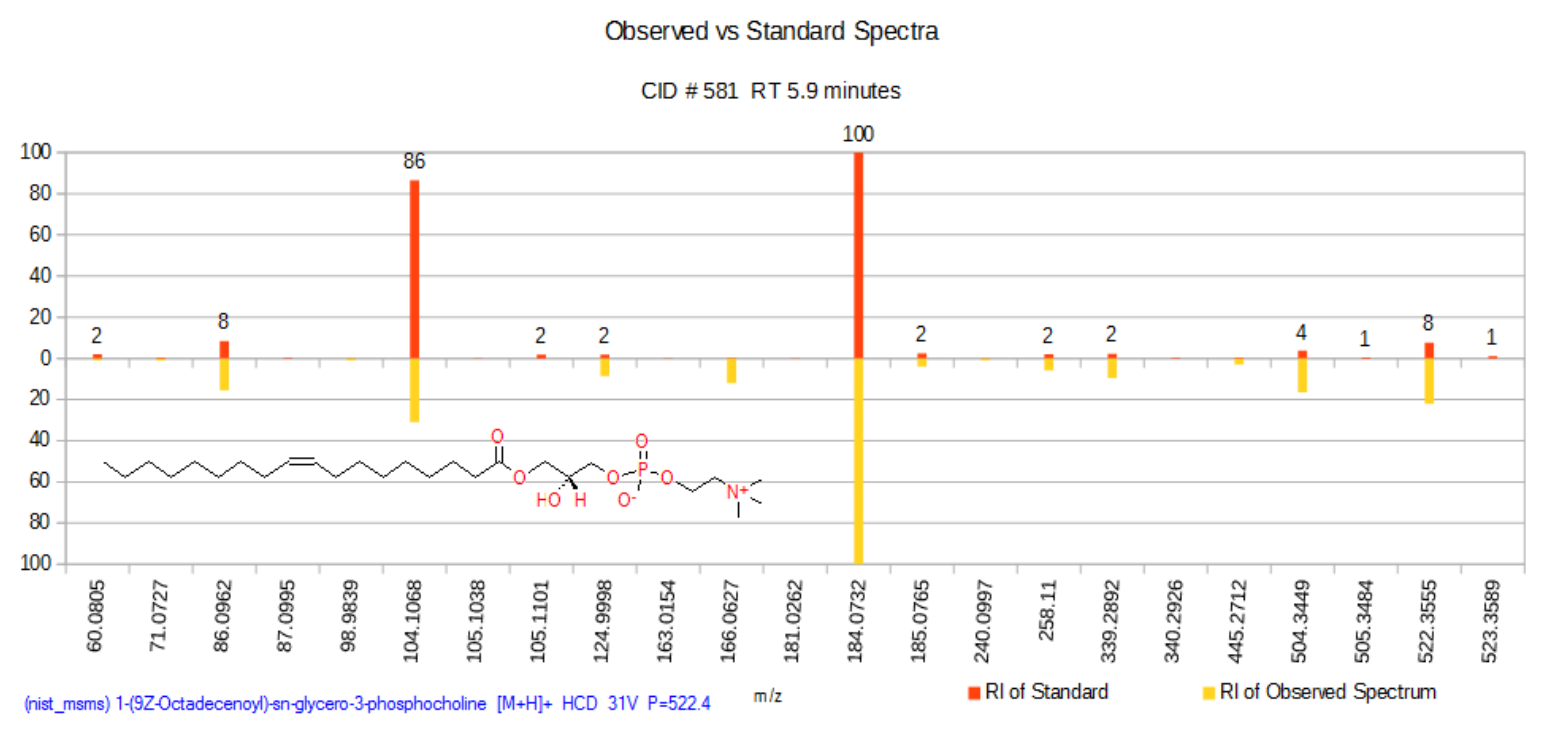
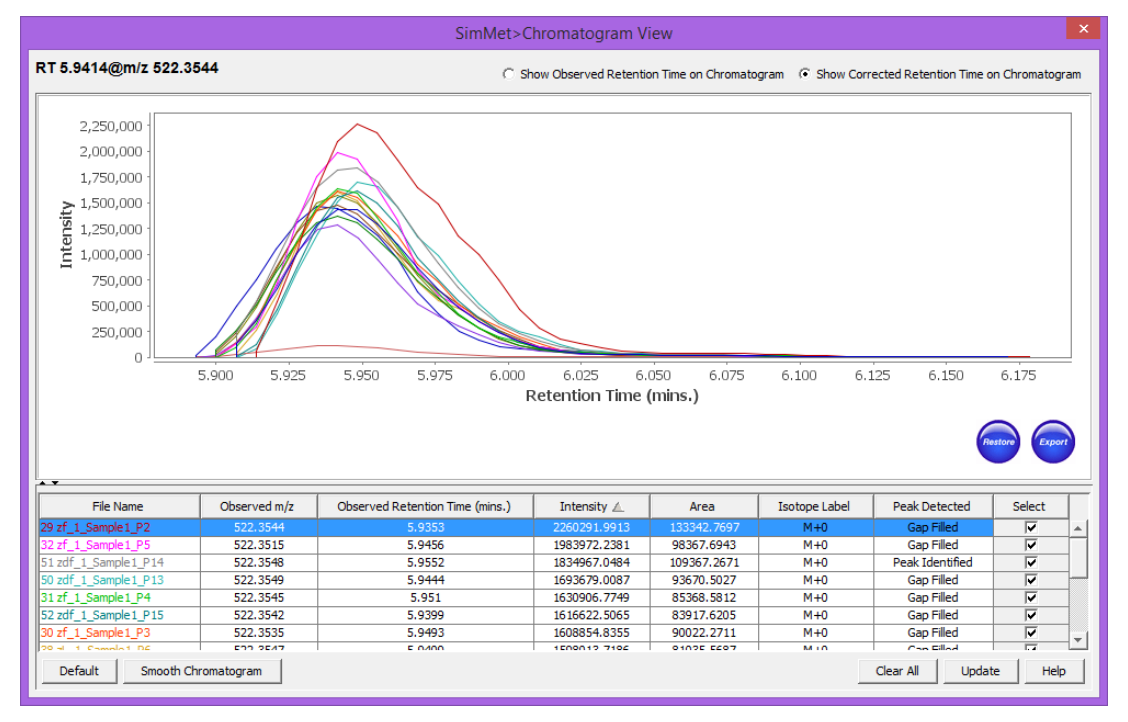


Figure 5(b): Typical SimMet software window for annotated MS/MS spectra of the identified metabolite



SimMet>Chromatogram View

RT 5.9414@m/z 522.3544

○ Show Observed Retention Time on Chromatogram ● Show Corrected Retention Time on Chromatogram

File Name	Observed m/z	Observed Retention Time (mins.)	Intensity	Area	Isotope Label	Peak Detected	Select
29 zf_1_Sample1_P2	522.3544	5.9353	2260291.9913	133342.7697	M+0	Gap Filled	☑
32 zf_1_Sample1_P5	522.3515	5.9456	1983972.2381	98367.6943	M+0	Gap Filled	☑
51 zdf_1_Sample1_P14	522.3548	5.9552	1834967.0484	109367.2671	M+0	Peak Identified	☑
50 zdf_1_Sample1_P13	522.3549	5.9444	1693679.0087	93670.5027	M+0	Gap Filled	☑
31 zf_1_Sample1_P4	522.3545	5.951	1630906.7749	85368.5812	M+0	Gap Filled	☑
52 zdf_1_Sample1_P15	522.3542	5.9399	1616622.5065	83917.6205	M+0	Gap Filled	☑
30 zf_1_Sample1_P3	522.3535	5.9493	1608854.8355	90022.2711	M+0	Gap Filled	☑

Figure 6: Typical SimMet software window showing aligned XICs of LysoPC 18:1 across samples

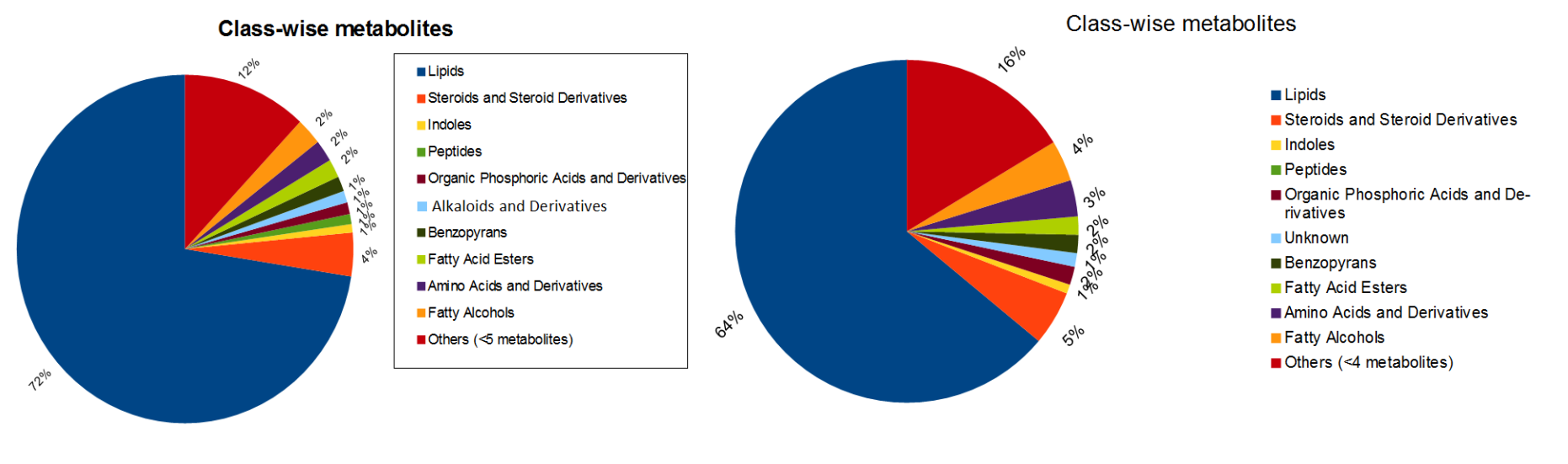


Figure 7: Pie chart showing breakdown of the metabolite profile on the basis of class (a): Total 643 (b): only metabolites with p<=0.05; Total 228

4

Principal Component Analysis: Classification of Samples

Data Set 1: All the LC-MS detected 2319 compounds including identified and unidentified compounds.
Data Normalization technique: Normalized to total intensity and Pareto scaling was used to remove intra-sample variation.
Key Observations: The first two principal components (PC1 and PC2) explained 47% of the total data variation. Score plot shows that the classification of samples is evident although some of the replicates are mixed up and one of the technical replicates of the sample "lean" is clearly an outlier. i.e., sample with very different properties from those within the ellipse. Samples close to each other have similar properties (Figure 8, right window).

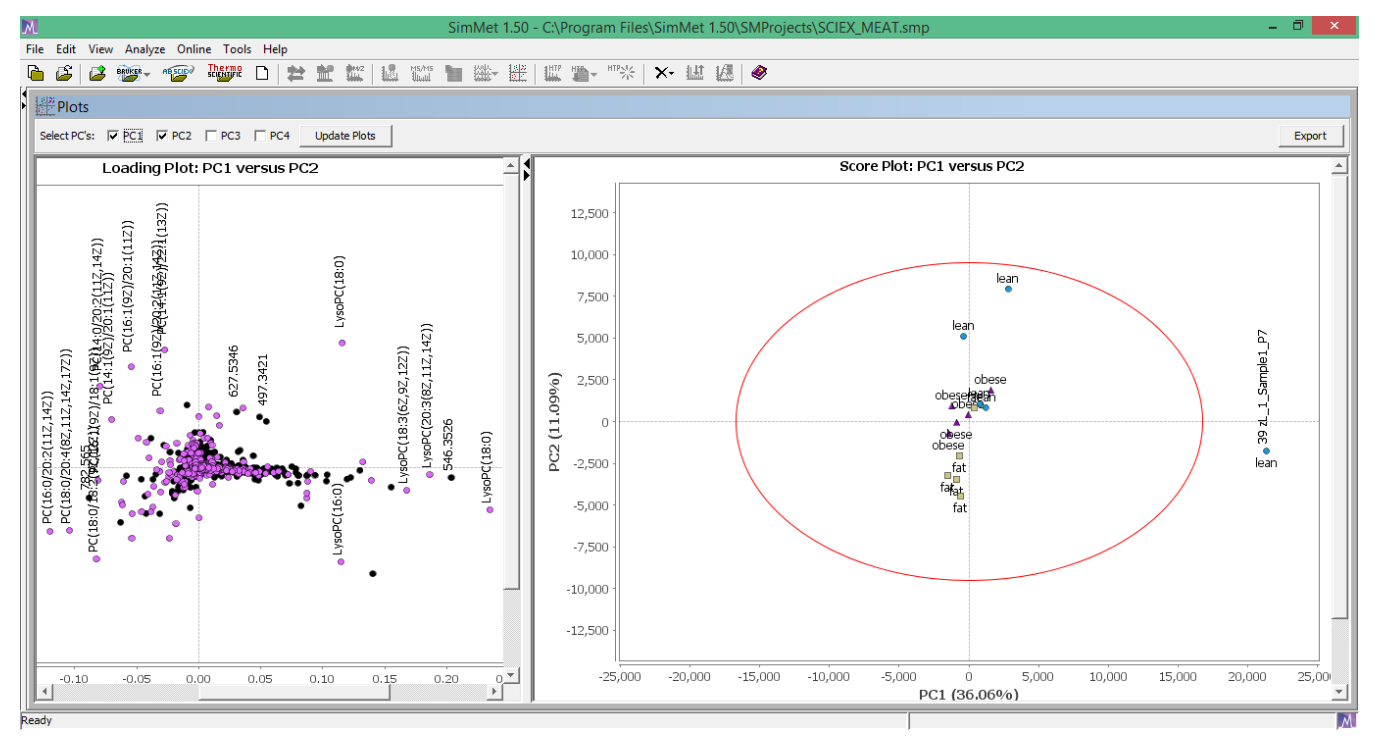

SimMet 1.50 - C:\Program Files\SimMet 1.50\SMProjects\SCIEX_MEAT.smp
Plots
Select PC's: ☑ PC1 ☑ PC2 ☐ PC3 ☐ PC4
Update Plots
Export
Loading Plot: PC1 versus PC2
Score Plot: PC1 versus PC2
lean
obese
fat
39 zL_1_Sample1_P7
PC2 (11.09%)
PC1 (36.06%)

Figure 8: SimMet software interface showing PCA loadings and score plots. One of the technical replicates of the sample "lean" is an outlier. Black dots in the loadings plot represents unidentified compounds.

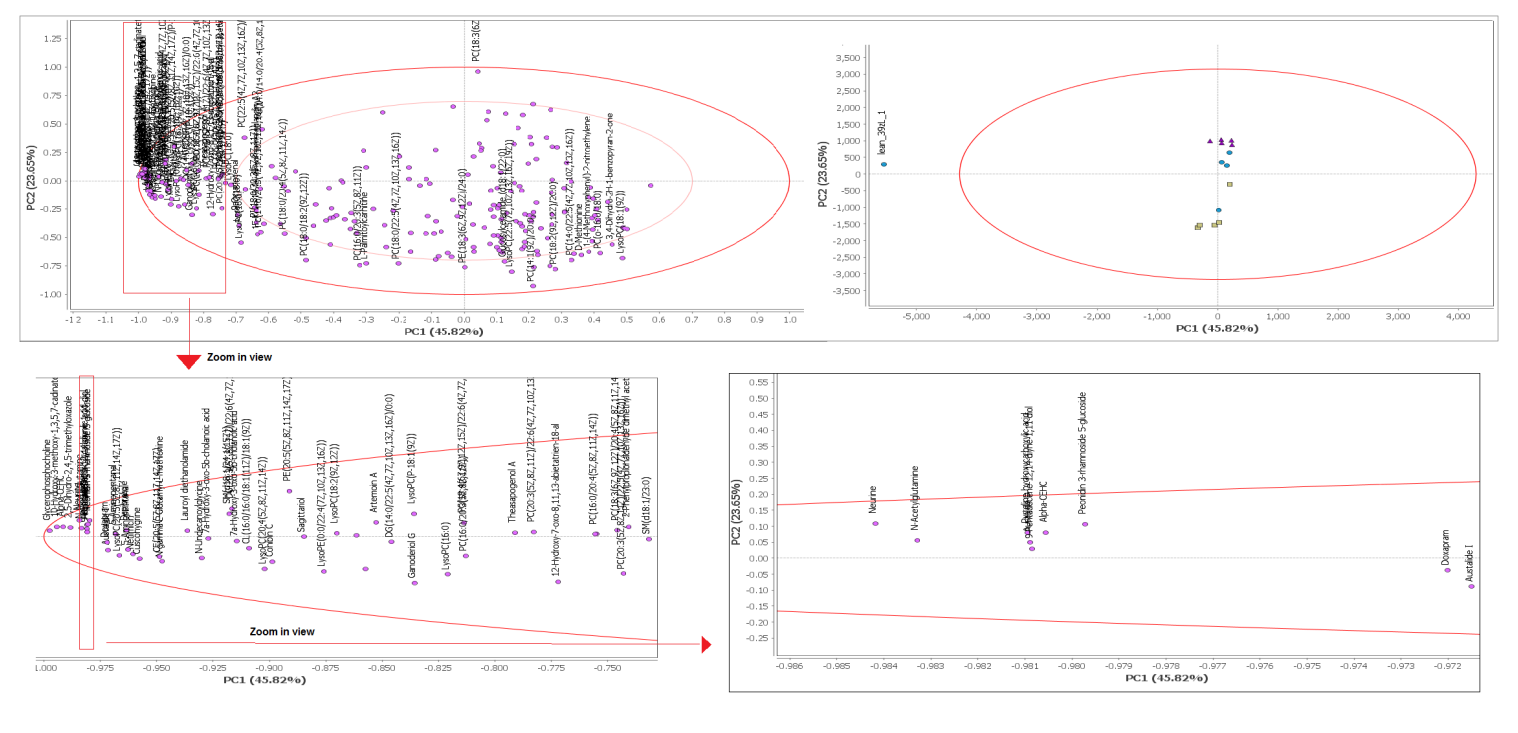
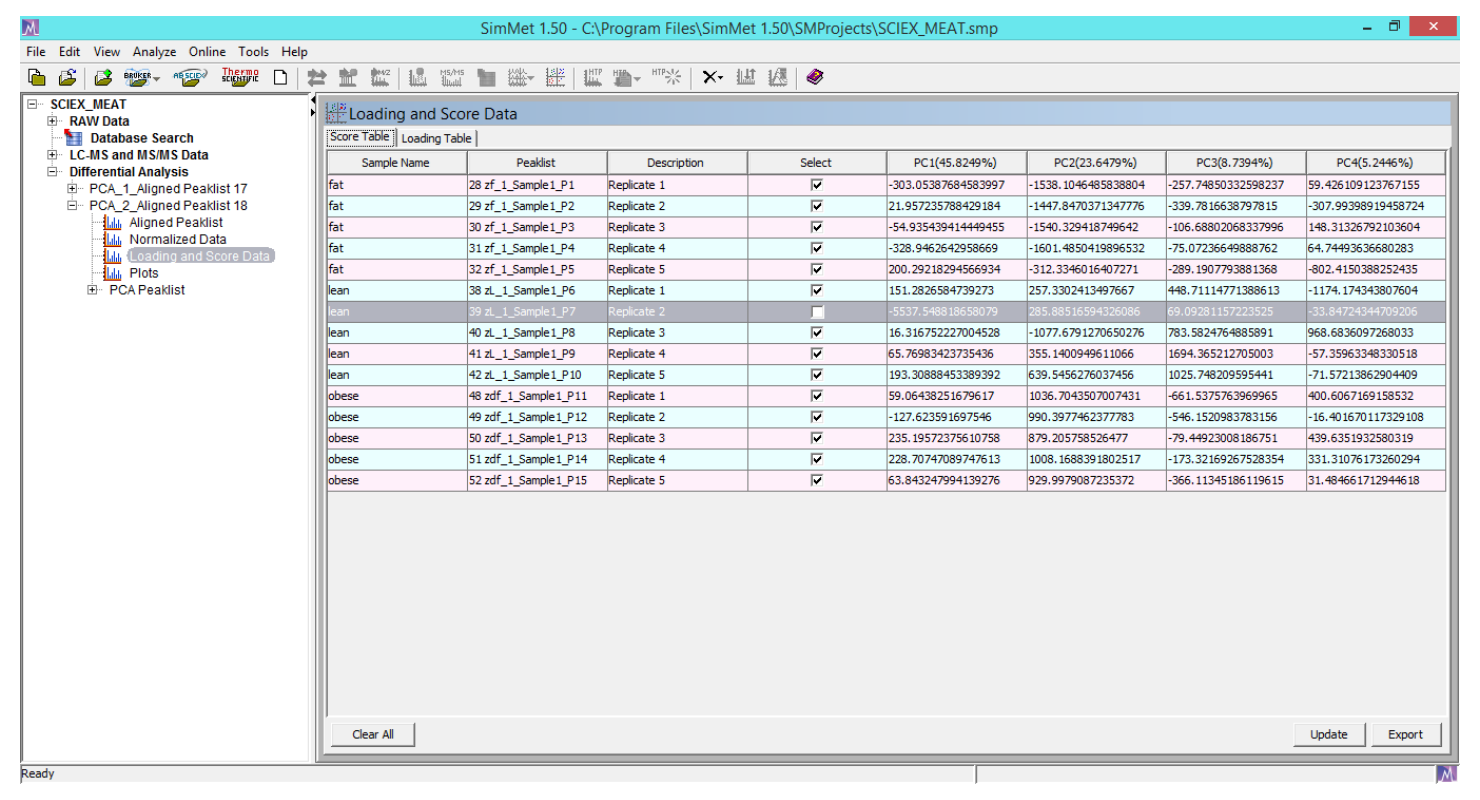


Figure 9: Typical SimMet software interface showing loadings and score plots from PCA performed considering only metabolites with p<= 0.05. The encircled area in the loading plot is zoomed in for better review of the metabolites.

Data Set 2: In order to achieve better classification of the samples, we create a subset of the data containing metabolites with p-value <= 0.05 i.e., 228 cases.
Key Observations: The data variation explained by PC1 and PC2 is increased (69.5%). However, the score plot in figure 9 still shows the outlier replicate of the sample "lean". This may call for further investigation of the particular technical replicate.
Data Set 3: We update the PCA by removing the replicate from the analysis (Figure 10). This leads to 10% reduction in the total explained data variation by PC1 and PC2. Despite this reduction, the classification of the biological samples is enhanced (Figure 11). We observe visible classification of the LC-MS runs on the basis of the samples namely lean, fat and obese.



SimMet 1.50 - C:\Program Files\SimMet 1.50\SMProjects\SCIEX_MEAT.smp

SCIEX_MEAT
- RAW Data
- Database Search
- LC-MS and MS/MS Data
- Differential Analysis
  - PCA_1_Aligned Peaklist 17
  - PCA_2_Aligned Peaklist 18
    - Aligned Peaklist
    - Normalized Data
    - Loading and Score Data
    - Plots
  - PCA Peaklist

Loading and Score Data (Score Table | Loading Table)

Sample Name	Peaklist	Description	Select	PC1(45.8249%)	PC2(23.6479%)	PC3(8.7394%)	PC4(5.2446%)
fat	28 zf_1_Sample1_P1	Replicate 1	☑	-303.05387684583997	-1538.1046485838804	-257.74850332598237	59.426109123767155
fat	29 zf_1_Sample1_P2	Replicate 2	☑	21.957235788429184	-1447.8470371347776	-339.781663797815	-307.99398919458724
fat	30 zf_1_Sample1_P3	Replicate 3	☑	-54.935439414449455	-1540.329418749642	-106.68802068337996	148.31326792103604
fat	31 zf_1_Sample1_P4	Replicate 4	☑	-328.9462642958669	-1601.4850419896532	-75.07236649888762	64.74493636680283
fat	32 zf_1_Sample1_P5	Replicate 5	☑	200.29218294566934	-312.3346016407271	-289.1907793881368	-802.4150388252435
lean	38 zL_1_Sample1_P6	Replicate 1	☑	151.2826584739273	257.3302413497667	448.71114771388613	-1174.174343807604
lean	39 zL_1_Sample1_P7	Replicate 2	☐	-5537.548818658079	285.88516594326086	69.09281157223525	-33.84724344709206
lean	40 zL_1_Sample1_P8	Replicate 3	☑	16.316752227004528	-1077.6791270650276	783.5824764885891	968.6836097268033
lean	41 zL_1_Sample1_P9	Replicate 4	☑	65.76983423735436	355.1400949611066	1694.365212705003	-57.35963348330518
lean	42 zL_1_Sample1_P10	Replicate 5	☑	193.30888453389392	639.5456276037456	1025.748209595441	-71.57213862904409
obese	48 zdf_1_Sample1_P11	Replicate 1	☑	59.06438251679617	1036.7043507007431	-661.5375763969965	400.6067169158532
obese	49 zdf_1_Sample1_P12	Replicate 2	☑	-127.623591697546	990.3977462377783	-546.1520983783156	-16.401670117329108
obese	50 zdf_1_Sample1_P13	Replicate 3	☑	235.19572376510758	879.205758526477	-79.44923008186751	439.6351932580319
obese	51 zdf_1_Sample1_P14	Replicate 4	☑	228.70747089747613	1008.1688391802517	-173.32169267528354	331.31076173260294
obese	52 zdf_1_Sample1_P15	Replicate 5	☑	63.843247994139276	929.9979087235372	-366.11345186119615	31.484661712944618

Figure 10: SimMet software interface facilitating removal of a technical replicate from the analysis and update the results.

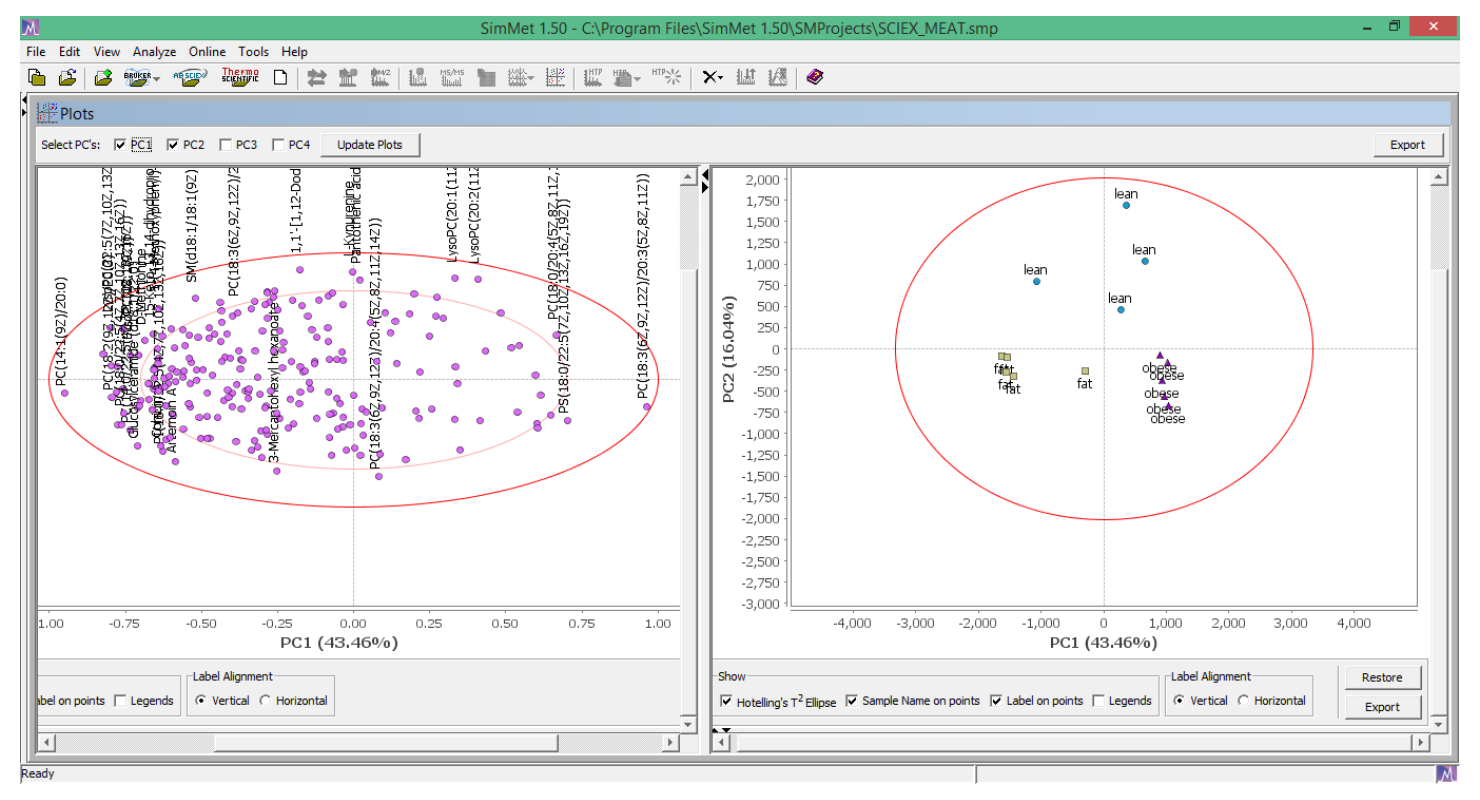

SimMet 1.50 - C:\Program Files\SimMet 1.50\SMProjects\SCIEX_MEAT.smp
Plots
Select PC's: ☑ PC1 ☑ PC2 ☐ PC3 ☐ PC4
Update Plots
Export
PC1 (43.46%)
PC2 (16.04%)
lean
fat
obese
Show: ☑ Hotelling's T² Ellipse ☑ Sample Name on points ☑ Label on points ☐ Legends
Label Alignment: ● Vertical ○ Horizontal
Restore
Export

Figure 11: PCA loadings and score plots after removing outlier.

5

Sl #	(i) Class	(ii) # metabolites	(iii) p <=0.05	(iv) PCA of (ii)*	(v) CA of 2)**		Sl #	(i) Class	(ii) # metabolites	(iii)p <=0.05	(iv) PCA of (ii)*	(v) PCA of (2)**
1	Glycerophospholipids	338	110	39	24		33	Glycosyl Compounds	2	1		
2	Sphingolipids	63	15	5	4		34	Diazines	2	1		
3	Fatty Alcohols	16	9	3	3		35	Benzyl Alcohols and Derivatives	2	1		
4	Amino Acids and Derivatives	13	8	3	3		36	Pyrrolopyridines	1	1		
5	Fatty Acid Esters	11	4	1	3		37	Phenylpropenes	1	1		
6	Glycerolipids	35	11	3	2		38	Nitrobenzenes	1	1		
7	Benzopyrans	9	4	3	2		39	Cinnamaldehydes	1	1		
8	Alkylamines	4	3	2	2		40	Benzofurans	1	1		
9	Prenol Lipids	30	13	6	1		41	Sugar Acids and Derivatives	2			
10	Steroids and Steroid Derivatives	26	12	4	1		42	Lineolic Acids and Derivatives	2			
11	Alkaloids and Derivatives	7	3	3	1		43	Keto-Acids and Derivatives	2			
12	Styrenes	2	1	1	1		44	Trichothecenes	1			
13	Amino Acids	1	1	1	1		45	Triazines	1			
14	Eicosanoids	1	1	1	1		46	Trialkylamines	1			
15	Disaccharides	1	1		1		47	Thiols	1			
16	Ethers	1	1		1		48	Thiazolidines	1			
17	Organic Phosphoric Acids and Derivatives	7	4	1			49	Tetralins	1			
18	Acenes	3	1	1			50	Pyrrolidines	1			
19	Flavonoids	2	2	1			51	Pyridines and Derivatives	1			
20	Diphenylmethanes	2	1	1			52	Piperidines	1			
21	Fatty Amides	2	1	1			53	Naphthopyrans	1			
22	Amino Sugars	1	1	1			54	Morphinans	1			
23	Fatty Aldehydes	1	1	1			55	Monosaccharides	1			
24	Oxazolines	1	1	1			56	Fatty Esters	1			
25	Phenylpiperazines	1	1	1			57	Fatty Alcohol Esters	1			
26	Pyrrolines	1	1	1			58	Dihydrothiophenes	1			
27	Indoles	5	2				59	Cinnamic Acid Derivatives	1			
28	Phenols and Derivatives	3	2				60	Carboxylic Acids and Derivatives	1			
29	Indanes	2	2				61	Biphenols	1			
30	Peptides	6	1				62	Benzoic Acid and Derivatives	1			
31	Fatty Acids and Conjugates	3	1				63	Acetophenones	1			
32	Pyrimidine Nucleotides	2	1									
								Total	637	228	85	51

Table 2. Summary of the LC-MS and MS/MS metabolite profiles reported as a result of different statistical analysis. *metabolites outside inner ellipse in Score Plot; ** PCA performed by removing 2nd technical replicate of the sample 'lean', and reporting the number of metabolites outside inner ellipse in score plot.

6

Differential Metabolites

As we have achieved desired classification of samples using the obtained metabolite profile, we further investigate which metabolites are influential and how the metabolites are correlated. In the loadings plots, metabolites that are located in the inner circle do not contain enough structured variation to be discriminating for the samples under investigation namely lean, fat and obese, i.e., explaining less than 50% of the variance in the data. The change in the number of effective metabolites from different classes in classifying the samples are displayed in figure 12 (a). The statistics is further changed after the removal of the outlier LC-MS run corresponding to 2nd technical replicate of the sample "lean" from the PCA. One key observation is that metabolites from 2 different classes become effective in classifying the samples which were not effective when the outlier replicate was considered in the PCA model. Further, two additional metabolites from class "fatty acid ester" becomes effective as compared to only one in the model with the outlier (Figure 13).

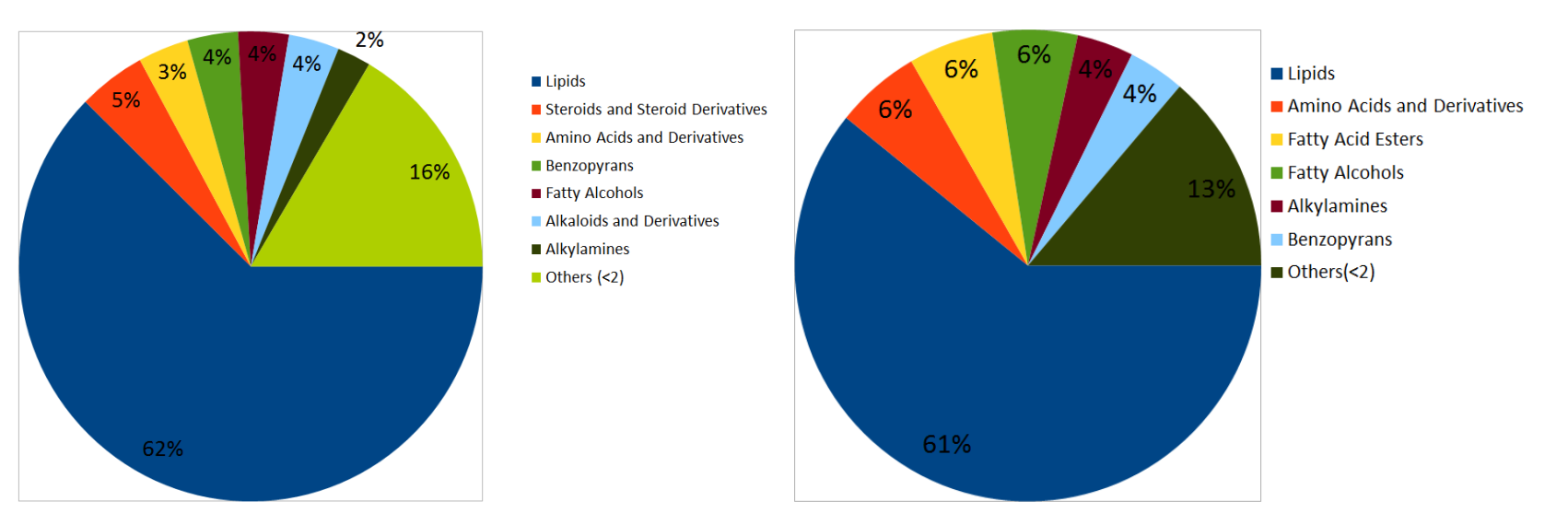


Figure 12: Pie chart showing breakdown of the metabolite classes that explain less than 50% of the samples variation. Data obtained on performing PCA of (a): metabolites with p-value<= 0.05; Total 85 (b): metabolites of (a) and having removed 2nd technical replicate of the sample "lean"; Total 51.

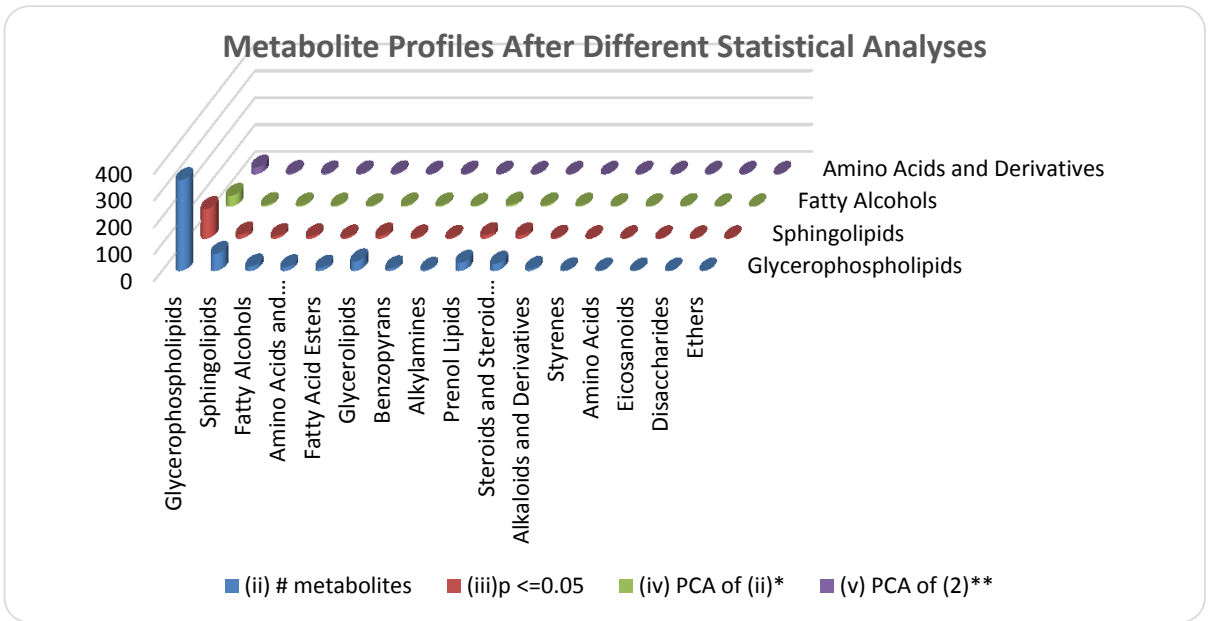


Figure 13: Change in the metabolite profile that have significant impact on classifying samples. Notice the increase in the number of effective metabolites from classes namely fatty acid esters, disaccharides and ethers in classifying the biological samples after removing the outlier LC-MS run from the data set.

Conclusion

SCIEX's TripleTOF® systems deliver high resolution, accurate mass data with high acquisition rates with the capability of acquiring MS and MS/MS information for as many features as possible providing a comprehensive interrogation of the samples. The systems offer one of the best suitable workflows for metabolite profiling. SimMet (PREMIER Biosoft, www.premierbiosoft.com), a standalone software tool supporting LC-MS data processing, metabolite identification, statistical analysis, and data visualization provides comprehensive data analysis and reviewing of the results in a single software platform. The software, with its proprietary algorithm, identifies metabolites by matching observed MS/MS data against consensus MS/MS spectra of standard metabolites collected for multiple experimental settings (NIST MS/MS database). This reliable proposal of compound identities helped save analysis time and money spent for purchasing multiple references in order to confirm the identity of the target compounds.

Acknowledgement: PREMIER Biosoft is a bonafide distributor of NIST MS/MS and NIST MS/MS2 databases. On procuring a license of SimMet, user will also receive a license of the mentioned NIST databases.

Reference

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. Gale D, and Lloyd S. S. 1962. American mathematical monthly : 9-15.
5. Fischler MA, Bolles RC. 1981; Comm Of the ACM, 24:381-395.
6. Cleveland WS, Devlin SJ. 1988; J Am Stat Assoc, 83(403):596-610.
7. Nordstrom, A. et al. 2006; Anal. Chem., 78, 3289-3295.
8. Voss, Bjorn, et al. (2011) Bioinformatics 27.7: 987-993.
9. Pluskal T et al, 2010; BMC Bioinformatics, 11:395