High resolution mass spectrometric data analysis using Progenesis QI software for non-targeted screening (NTS)

Stefania Della Gatta, Christoph Buchholz, Mark Bentley and Daniel Arndt

Philip Morris International R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland (part of Philip Morris International group of companies)

stefania.dellagatta@pmi.com

Introduction

Non-targeted screening (NTS) is a key methodology for characterizing the chemical composition of complex matrices (e.g. metabolomics samples) using an unbiased approach. Even with the use of liquid chromatography coupled to high resolution accurate mass spectrometry (LC-HRAM-MS), compound identification and subsequent structural elucidation can be time consuming and highly challenging. Here we present a novel NTS strategy that combines both full scan and data-dependent fragmentation high resolution accurate mass spectrometry and data processing performed using metabolomics software (Progenesis QI™) in combination with experimental MS^2 fragmentation databases. A semi-automated comparison of measured fragmentation spectra versus *in-silico* predicted fragmentation for putative compound hits was also used to enhance the characterization of complex matrices. Therefore, data evaluation time for complex data sets could be dramatically reduced with enhanced confidence in compound identification.

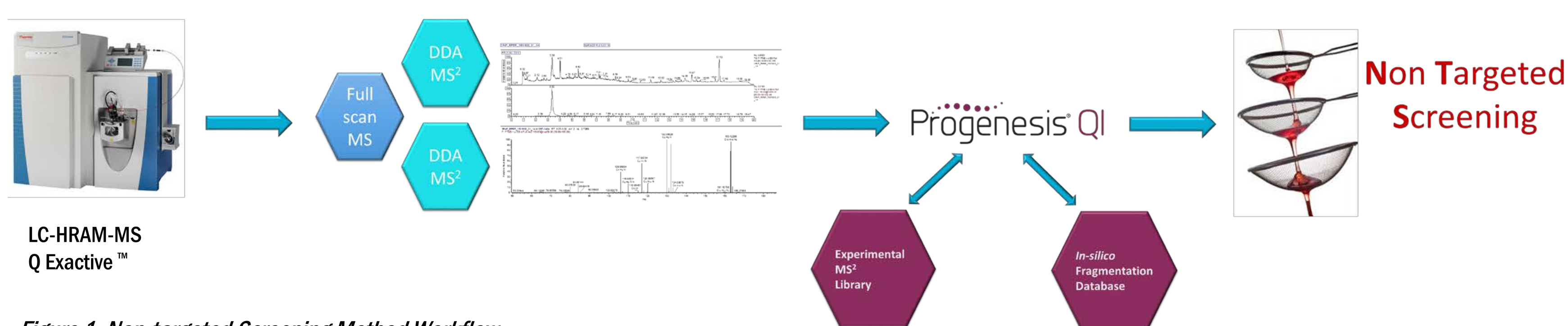

Figure 1. Non-targeted Screening Method Workflow

Method

Data generation was performed in full scan mode combined with high-energy collision dissociation (HCD) and stepped normalized collision energy (NCE) applied using a Q Exactive™ high resolution accurate mass spectrometer (Thermo Fischer Scientific, Bremen, Germany). Sample replicates (*n* = 5) of 2 or more samples, a blank and a pool sample (equal proportions of each sample including the blank in order to combine all compound features within a single reference sample) were fortified with a set of stable isotope labeled internal standards to enable semi-quantification. Analysis was performed using reversed phase (RP) chromatography with positive and negative heated electrospray ionization (HESI+/HESI-) and positive atmospheric pressure chemical ionization (APCI+) to cover a wide range of substances with different ionization properties. In addition, samples were analyzed using hydrophilic interaction chromatography (HILIC) with HESI+.

Table 1. Gradient RP Modes

Time [min]	A [%]	B [%]
0	85	15
7.00	10	90
12.80	0	100
18.00	0	100
18.10	85	15
20.00	85	15

Column: Hypersil GOLD™ (150 × 2.1 mm, 1.9 µm)
Guard: UHPLC filter cartridge (10 × 2.1 mm, 0.2 µm)
Flow [µL/min]: 400
Temperature: 50°C
HESI+, APCI+ mode:
Mobile phase A: 10mM NH_4Ac in water,
Mobile phase B: 1mM NH_4Ac in methanol
HESI- mode:
Mobile phase A: 1mM NH_4F in water
Mobile phase B: methanol

Table 2. Gradient HILIC Mode

Time [min]	A [%]	B [%]
0	2	98
7.00	25	75
8.00	2	98
15.00	2	98

Column: Accucore HILIC™ (150 x 2.1 mm, 2.6 µm)
Guard: Defender guard HILIC cartridge (10 × 2.1 mm, 2.6 µm)
Flow [µL/min]: 500
Temperature: 50°C
HESI+ mode:
Mobile phase A: 10mM NH_4Ac in water
Mobile phase B: 10mM NH_4Ac in acetonitrile

HRAM Detection

Full scan MS was performed at a resolution of 70.000 (FWHM) acquiring a mass range of m/z 80 – 800 in combination with a data-dependent MS^2 Top3 of each scan at a resolution of 17.500 (FWHM) and applied stepped normalized collision energies of 25, 50 and 75 eV and automated gain control of 1 x $10e^5$ in order to generate HCD first order fragmentation (TopN = 3, loop count = 3, dynamic exclusion = 10 s). Vaporizer heater temperature, capillary heater temperature, spray voltage, sheath gas and auxiliary gas were set at 350˚C, 380˚C, ±3.00 kV, 60 and 20 arbitrary units respectively for HESI modes. Vaporizer heater temperature, capillary heater temperature, discharge current, sheath gas and auxiliary gas were set at 450˚C, 380˚C, 5.0 µA, 50 and 5 arbitrary units respectively for APCI mode.

Data Processing

Acquired data were processed using Progenesis QI ™ software (Nonlinear Dynamics, Newcastle upon Tyne, UK), consisting of raw data import, selection of possible adducts, peak set alignment, peak detection, deconvolution, data set filtering, noise reduction, compound identification and normalization with internal standards. In the developed workflow, further data evaluation steps are performed for compound identification.

Fundamental steps within the Progenesis QI™ data evaluation workflow:

- Selection of possible adducts
  - $[M+H]^+$, $[M+NH_4]^+$ for RP HESI+ and HILIC HESI+
  - $[M+H]^+$, $[M+H-H_2O]^+$ for RP APCI+
  - $[M-H]^-$, $[M+F-H]^-$ for RP HESI-
- Importing of Thermo.raw profile data
- Visual quality check of each analytical run using ion intensity maps
- Alignment with a selected reference run (one of the pool samples)
- Experimental design setup (defining one or more groups for aligned runs)
- Peak picking
- Normalization versus a set of internal standards
- Automatic deconvolution to enable accurate quantification of each compound

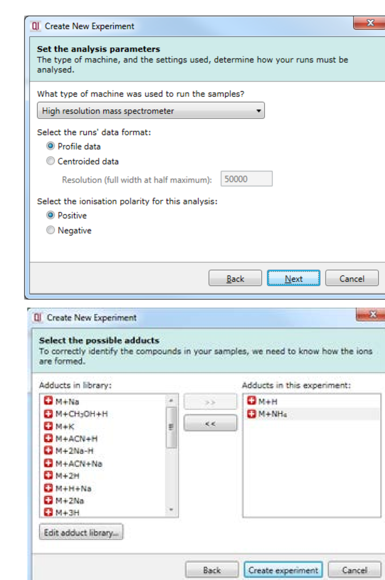

Figure 3. Create new Experiment sections in Progenesis QI™

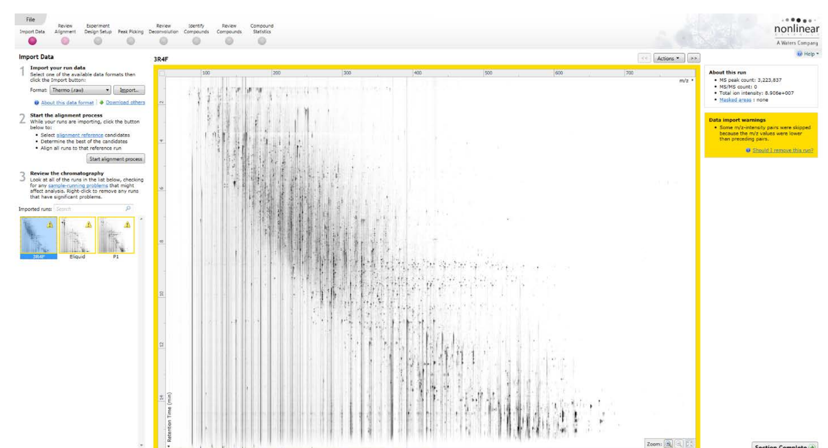

Figure 4. Data Import Window in Progenesis QI™

Figure 5. Alignment Window in Progenesis QI™

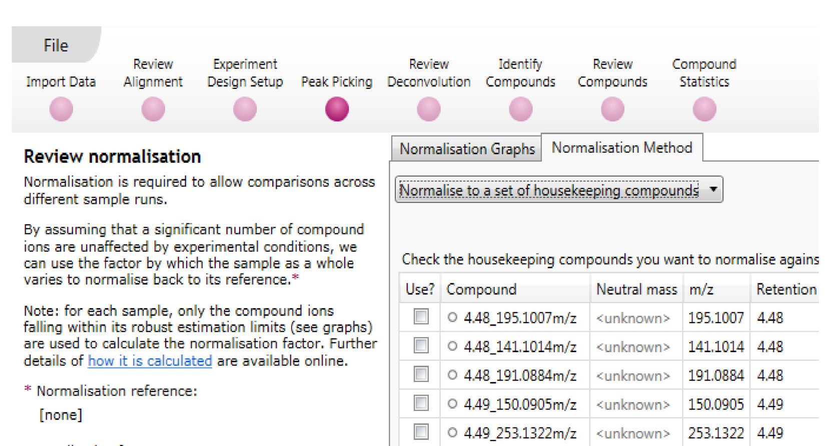

Figure 6. Normalization Window in Progenesis QI™

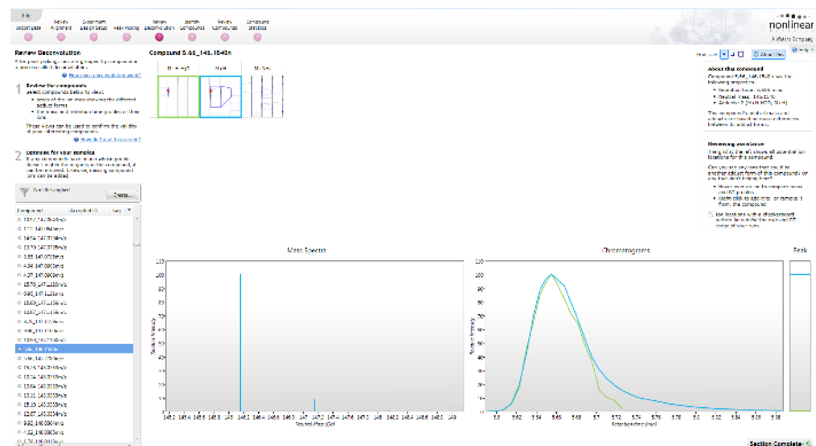

Figure 7. Review Deconvolution Window in Progenesis QI™

Results

- Deconvolution Review

In order to increase the confidence for semi-quantification and compound identification all detected ions, including various adducts, for a compound were automatically combined. Mass spectra and extracted ions were visualized to assess how similar they are. The assignment of the correct adduct was essential for finding a good compound candidate hit. The deconvoluted compound spectra were reviewed to assess if their profiles appeared as an outlier in terms of m/z and retention time characteristics, outside expected limits, and could be removed if necessary.

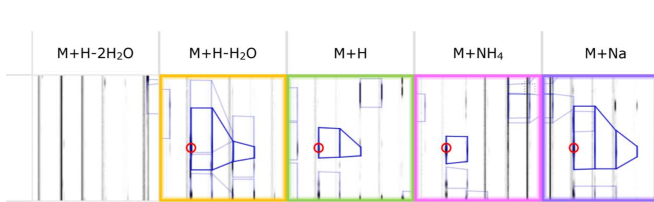

Figure 8. Montage view showing location of detected adducts for a compound

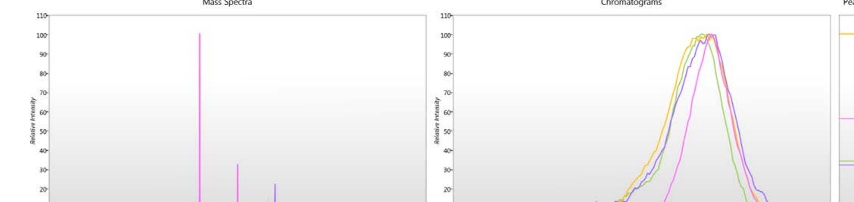

Figure 9. Mass spectra and chromatograms for the detected adducts of a compound, color coded by adduct

- Compound Identification

A revolutionary and major part of the workflow for dealing with high resolution accurate mass first order fragmentation data is the semi-automated process for compound identification. Compound identification was performed using a stepwise approach employing experimental MS^2 fragmentation databases and *in-silico* predicted fragmentation of chemicals from public databases. In Step 1 all detected constituents were matched and assigned against an in-house database comprising experimental data for approximately 400 reference compounds with accurate mass data, stepped NCE MS^2 first order fragmentation and retention times (precursor and fragment tolerance 5ppm, retention time tolerance 0.5 min). In Step 2 fragmentation patterns for all detected constituents were compared with *in-silico* predicted fragmentation of putative hits from UCSD (Unique Compounds & Spectra Database, PMI, Neuchâtel, Switzerland)[1], HMDB 3.6 (Human Metabolome Database, University of Alberta, Edmonton, Canada)[2,3,4] and, via the ChemSpider search plugin, with ChemIDplus (ChemIDplus, SIS, NLM, NIH, Bethesda, MD, USA) and FDA (U.S. Food and Drug Administration, Silver Spring, MD, USA) (precursor and fragment tolerance 5ppm). In Step 3 fragmentation spectra for detected constituents were compared with experimental fragmentation spectra of NIST14 MS/MS library (precursor and fragment tolerance 5ppm) (U.S. National Institute of Standards and Technology, Gaithersburg, MD, USA). All putative hits were scored using Progenesis QI™ algorithms, which considered mass similarity, isotope similarity and fragmentation score.

Figure 10. Experimental MS^2 Database Comparison

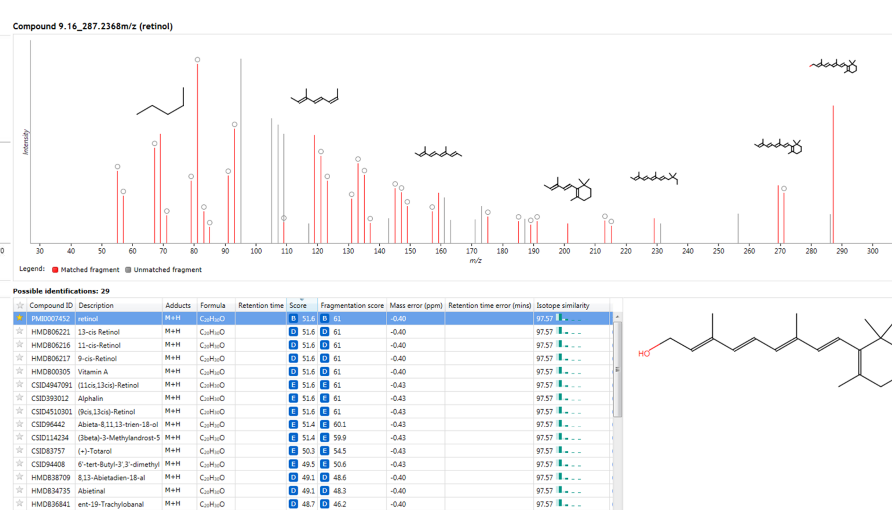

Figure 10. in-silico Fragmentation Database Comparison

- Compound Review

A manual review process was performed to ensure the correctness of compound identification. Each putative hit was checked regarding compound abundance, detected adducts, fragmentation score, retention time score, isotope similarity, mass error and overall score. This review step also considered the measured isotopic distribution compared to theoretical. This isotopic distribution match contributed to the overall compound identification score.

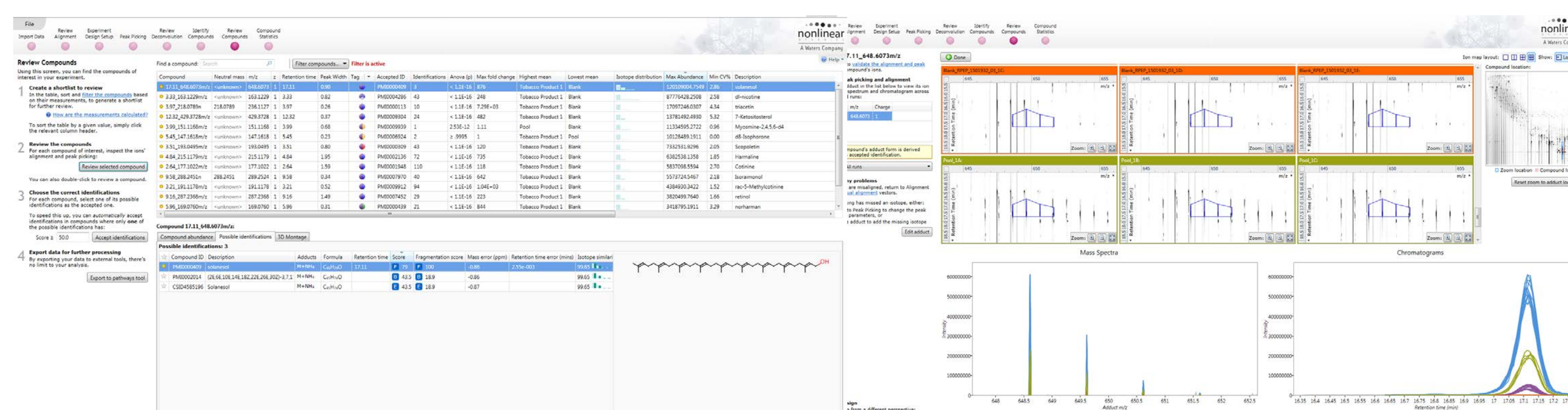

Figure 11. Compound Review Window in Progenesis QI™

In the 'Review Compounds' stage the behavior of compound subsets can be examined based on tag filters. Good power for differentiating single compounds between sample groups has been shown and basic statistical evaluations were performed.

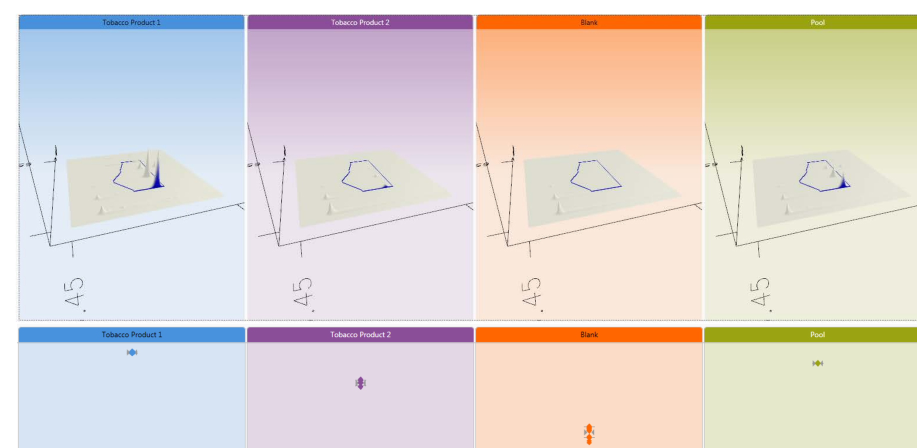

Figure 12. Compound abundance plot in Progenesis QI™

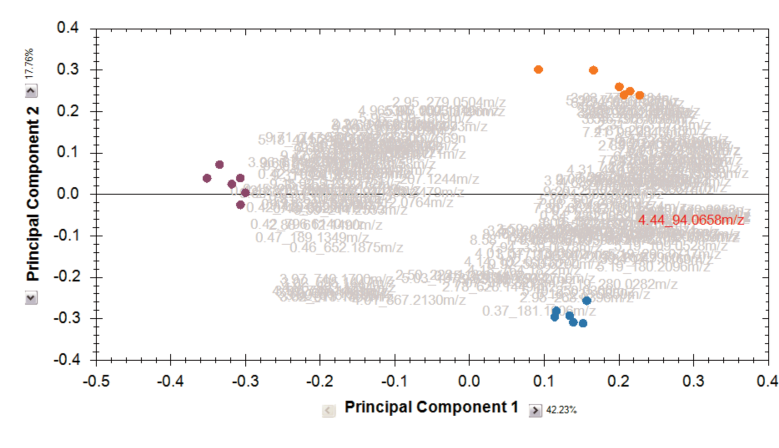

Figure 13. Principal Component Analysis in Progenesis QI™

Definitive compound confirmation was performed using reference standards matched with experimental first order fragmentation, isotopic similarity and retention time.

Conclusions

This non-targeted screening workflow using LC-HRAM-MS, in combination with Progenesis QI™ software, has been demonstrated to be a powerful tool that provides large-scale qualitative and semi-quantitative analysis of complex datasets with increased processing speed and improved confidence in compound identification for complex matrix characterization. The use of full scan accurate mass data in combination with first order fragmentation spectra enables a robust and efficient process for identifying detected unknown compounds.

References

1.Martin E., Monge A, et al., Building an R&D chemical registration system, Journal of Cheminformatics 2012 4:11, DOI: 10.1186/1758-2946-4-11
2.Wishart DS, Tzur D, Knox C, et al., HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007 Jan;35(Database issue):D521-6
3.Wishart DS, Knox C, Guo AC, et al., HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 2009 37(Database issue):D603-610.
4.Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, et al., HMDB 3.0 — The Human Metabolome Database in 2013. Nucleic Acids Res. 2013. Jan 1;41(D1):D801-7.

PMI SCIENCE
PHILIP MORRIS INTERNATIONAL
www.pmiscience.com

NonTarget 2016, Ascona, Switzerlandand

May 29 - June 3, 2016

Competing Financial Interest

The research described in this poster was funded by Philip Morris International