High-Throughput Qualitative and Quantitative Analysis of High Resolution Accurate Mass Data for Early Drug Discovery

Lester Taylor[1], Karen Salomon[1], Caroline Ding[1], Hans Pfaff[2], David Wright[1], Jim Shofstahl[1], Prakash Chand[1] and Zesmaa Morimoto[1]

1.Thermo Fisher Scientific, San Jose, CA, USA 2. Thermo Fisher Scientific, Bremen, Germany

Overview

Purpose: We describe the use of instrument software for metabolism screening which controls acquisition and processing of ultra-high resolution accurate mass data from the Thermo Scientific Exactive LC-MS instrument.
Methods: The software incorporates several novel algorithms for data processing, including advanced component detection, parameter-less peak integration and intelligent elemental composition determination.
Results: The software can be used to screen and report the nature and extent of parent drug metabolism in an automated fashion. At the same time, information is obtained for putative metabolites and these results are also output.

Introduction

The availability of full scan ultra high resolution LC-MS instruments allow the analysis of samples for drug discovery, metabolism and pharmacokinetics (PK) without the need to set up specific methods for target compounds. After acquisition, significant data processing is required to mine the data for parent drug and putative metabolites together with an assessment of quantitative and qualitative changes for time course experiments such as used in *in vitro* and *in vivo* compound assessment studies. We describe the use of software which will automatically process and directly report summary PK data.

Methods

In order to automatically process full scan ultra high resolution data it is necessary to employ software algorithms which will specifically manipulate data in order to identify compounds of interest at mass-to-charge values which have not necessarily been pre-defined. This allows the detection and reporting of qualitative and quantitative information for drug metabolites in time course studies in which a parent drug is specified. Components of the overall data processing methodology include: advanced background subtraction, and peak picking using parameter-less peak detection. Sophisticated elemental composition determination is used to identify parent drug related metabolites.

Workflow

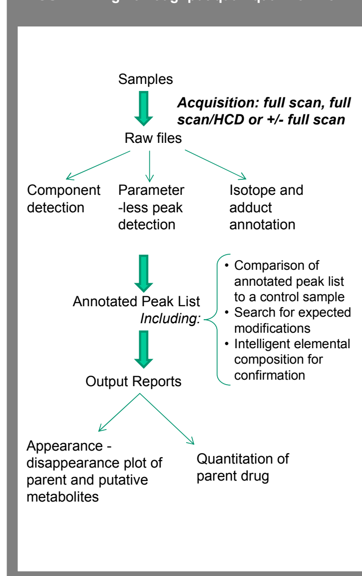
Prototype software has been developed to acquire and process high resolution, accurate mass metabolite data. A schematic of the workflow is shown in Figure 1. After the raw data files are acquired, a novel algorithm is used to detect all components having an acceptable chromatographic profile and a master peak list is created of all components detected. Peak area calculations are performed on these components using a parameter-less peak detection algorithm. Another algorithm is used to determine which components are isotopes or adducts of other components and these peaks are labeled accordingly in the master peak list.

The next step is to compare the peak list of one set of time points to a reference peak list – either a time zero data file or a control file. Any peaks common to both lists are labeled as such and are most likely not metabolites.

Metabolite finding is carried out on mass peaks in a sample that are different from the reference file. Metabolite finding will look at each un-annotated entry in the master peak list for known bio-transformations based on arithmetic difference. Peaks matching an expected modification will be labeled with that modification in the master peak list. Peaks not matching an expected modification remain un-annotated, but are not removed from the master peak list.

Plots of drug/metabolite concentration as a function of time are generated using the annotated peak lists for each time course sample. Relative quantitation of target compounds (parent drug and/or metabolites) is based on a user-defined reference. Confirmation of putative metabolites is provided with an elemental composition calculation.

FIGURE 1. High-throughput quan/qual workflow

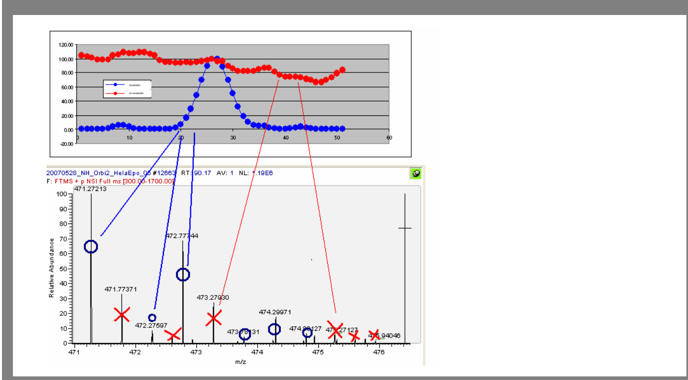


Automated Component Detection

The algorithm for advanced component detection is designed to remove noise and baseline signals from a data file. It can detect and separate components from each other even if their concentrations are different and their retention times are similar. Processing can be applied to chromatographic full scan high resolution MS spectra, i.e. LC-MS data files. There is a requirement that a minimum of three (or more) MS spectra should be acquired for each eluting component.

The algorithm performs a scan-by-scan analysis of the data file and searches for mass peaks within each scan. For each mass peak present, a mass chromatogram is created. The mass chromatogram of each peak within the scan is then evaluated to determine whether it has an acceptable chromatographic profile (Figure 2). For those masses with an acceptable profile, the centroid retention time is calculated and the mass is added to an event list with the centroid retention time. (The centroid retention time is used because it is more accurate.) The analysis is carried out for each scan in the data file to create a comprehensive event list of all components.

FIGURE 2. Examination of mass peaks in one scan at RT = 90.17 minutes. Mass peaks with acceptable chromatographic profiles are labeled with blue circles and mass peaks with un-acceptable chromatographic profiles are labeled with red crosses



The next step is to generate a peak list from the list of components or events. The peak list is created by grouping together all the events that have the same retention time within a tolerance window. The final peak list represents the reprocessed chromatogram as shown in Figure 3. The cleansed chromatogram offers improved component detection through the elimination of background signal.

FIGURE 3. Comparison of a reprocessed chromatogram to the total ion chromatogram. The top trace is the original chromatogram and the bottom trace is the cleansed chromatogram based on the peak list.

Annotated Peak Lists

Prototype software was used to analyze time course raw files for a metabolism sample. The sample files were processed to determine relevant components within each data file and to create the master peak list. The resulting peak list for each time course sample was then compared to a reference sample, in this case the sample at the zero time point. Components found only in the reference sample are flagged "control" in the master peak list of the sample file. Components that are unique to the time course sample were then searched against an expected modification database. If the component matched an expected modification, it was annotated as such (Figure 4). Peak area calculations for each putative metabolite were generated using the parameter-less peak detection algorithm and added to the master peak list.

FIGURE 4. Analysis of a tolbutamide (m/z = 271.1111 for MH^+) metabolism sample showing the master list annotations. Details of the software output are shown for the 6 hour time point. The other time points are accessed by the tabs.

Intelligent Elemental Composition Determination

Confirmation of putative metabolite assignment is performed using an intelligent elemental composition determination based on isotope accurate mass and intensity ratios (Figure 5). A single mass – usually the mono-isotopic mass – of a measured isotope pattern is used to calculate all possible elemental compositions that lie within a tolerance window that is determined from the expected mass error of the measurement. Then the theoretical isotope pattern for each elemental composition candidate is calculated. In the next step, the "Spectral Fit" between the theoretical and measured isotope pattern is calculated. The Spectral Fit is a number between 0% (where the patterns are completely different) and 100% (where the patterns are identical). The composition candidates are sorted in decreasing spectral fit order.

FIGURE 5. Intelligent Elemental Composition calculation (blue) compared to a traditional formula calculation (red) around one exact molecular mass. The Intelligent Elemental Composition algorithm takes into consideration isotopic masses and their intensities.

Time Course Study Results

Prototype software incorporating automated processing algorithms was used to process *in vitro* time course data from a study in which tolbutamide (10 µM) was incubated with rat hepatocytes. Time course samples were collected at 0, 2, 4, 6 and 24 hours and analyzed by LC-MS using the Exactive™ and Thermo Scientific Accela LC. Alternating full MS and HCD scans from m/z 85-850 were acquired.

The relative area of each putative metabolite is calculated as a ratio to the parent drug area at time = 0 for each time point. Relative areas of the putative metabolite were then plotted at the different time points to generate a graph showing the appearance of the metabolite. On the same graph, there is also a plot of the relative parent disappearance data. The plot provides comprehensive information on the metabolic stability of the target drug compound.

Figure 6 shows the metabolic stability plot of the parent and the hydroxylated metabolite. It can readily be seen that tolbutamide has metabolized very slowly in this system and that only one expected modification was observed. The formula calculations confirm the parent drug identity and the metabolite identity.

FIGURE 6. Metabolic stability plot of tolbutamide (m/z = 271.1105 for MH+) and oxidation metabolite in rat hepatocytes incubation. Insets are elemental composition calculations.

Conclusions

- We have demonstrated the use of a software-prototype that can automatically process and report summary PK results using full scan, high resolution, accurate mass data
- The incorporation of a novel algorithm allows detection of all components in a data file without prior knowledge of target compounds in the sample The software can analyze each data file in a time course experiment and generate a master peak list of all compounds found for each time point.
- Confirmation of the biotransformation formula is provided by an advanced elemental composition calculation.
- A summary plot of the time course behavior of the parent drug and the appearance of any metabolites found is generated from the master peak list. A visual inspection of the plot can rapidly determine the extent of parent drug metabolism and the identity of putative metabolites.
- Further work will be carried out to confirm robustness of the algorithm for low-level metabolites and for metabolites in complex matrices, and to evaluate the algorithms for absolute quantitation.

Acknowledgements

We would like to thank Yingying Huang, Kristi Akervik and Rose Herbold from Thermo Scientific, San Jose, CA for supplying the tolbutamide time course data.

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is not intended to encourage use of these products in any manners that might infringe the intellectual property rights of others.

IMSC09_M075_LTaylor