Waters

ADVANCES IN AUTOMATED METABOLITE IDENTIFICATION SOFTWARE TOOLS COUPLED WITH IN-SILICO METABOLITE PREDICTION USING LC-TOF-MS-MS

Andrew Baker[1*], Jennifer Granger[1], Jose Castro-Perez[2*], Jane Kirby[2], Michael McCullagh[2], Nicole McSweeney[3]

[1]Waters Corporation, Milford, Massachusetts, USA; [2]Waters Corp UK, Floats Road, Wythenshawe, Manchester, UK; [3]Lhasa Ltd, Dept of Chemistry, University of Leeds, Leeds, UK

ABSTRACT

Within the pharmaceutical and biotechnology industry, LC/MS is widely used for both quantitative and qualitative bioanalysis in different biological fluids in DMPK. Continuous improvements in LC/MS instrumentation have led to increased sample throughput capability. The major obstacle is the ability to process the data in order to extract the required information about metabolic route and rate, identify metabolic "soft spots" and identify compound or metabolic liabilities to support the "fail fast, fail cheap" paradigm of drug discovery. This is especially apparent when looking at low level metabolites from *in vitro* and *in vivo* screens, which are difficult to see in the TIC and require manual data manipulation to characterize these putative metabolites.

Software packages such as the Metabolynx™ Application Manager automate the routine and labour-intensive task of extracting potential metabolite peaks from complex chromatograms and present the results in a reduced data set for rapid review. A newly devised filter based on accurate mass has been incorporated into this data processing algorithm. Comparison with a matrix control sample eliminates non drug-related peaks and other endogenous components. Once the metabolites have been detected, exact mass and elemental composition data are passed on to an in-silico metabolite prediction knowledge base system, METEOR. This software program correlates the observed mass spectrometric data with predicted metabolic biotransformations based on the structure of the starting drug. As an additional filter, MS/MS may be used to help distinguish drug-related compounds from endogenous peaks. Several *in vitro* and *in vivo* examples will be shown to highlight differences in data processing throughput and information content using the improved algorithm and increased functionality when compared to manually processed data.

INTRODUCTION

The identification of drug metabolites in drug discovery and development is a difficult and time-consuming process. Traditionally the manual task of sifting through paper copies of multiple, complex data sets to confirm the presence of predicted biotransformations is very labour intensive. The use of LC/MS has become a standard analytical tool in metabolism laboratories over the last decade. However, unlike UV detection, the typical chromatographic trace acquired (the total ion chromatogram; TIC) often shows little obvious evidence of analytes amongst the background signal of a complex biological sample matrix. Each spectrum in the chromatographic time frame must be individually checked for evidence of new components.

This process, although time consuming, yields confirmation of expected metabolites, based on prior knowledge of the experienced metabolism scientist. However unexpected components are also common and are not so easily identified. For several years, the pharmaceutical industry has been very successful in applying to their drug metabolism studies hybrid quadrupole orthogonal time of flight mass spectrometers. In turn, this has allowed scientists to obtain exact mass data for both MS and MS/MS to identify metabolites with great confidence. The bottleneck is no longer in producing analytical data—it has shifted to the processing and interpretation of these data sets to extract useful information for decision-making.

A software tool has been previously described[1] that automatically processes LC/MS data sets to search for both expected and unexpected metabolites. This work demonstrates advances in the Metabolynx Application Manager that significantly improves performance of the data processing. The acquisition of exact mass data is the key to maximise the capability of the software to accurately identify real metabolites. Examples will be shown to demonstrate how a novel algorithm can automatically exclude components present in the control sample to yield an easily manageable number of entries in the analyte sample. Moreover, in order to minimise the number of false positives another new algorithm based on exact mass filter will be fully described in this paper. Once the data processed has been filtered with these algorithms then the elemental compositions for each one of the metabolites detected will be automatically passed to METEOR for in silico metabolite prediction and filtering from results of Metabolynx.

METABOLYNX™—HOW DOES IT WORK?

Metabolynx is a software application manager which automatically detects putative biotransformations for expected and unexpected metabolites. The application manager automatically runs samples scheduled for analysis by LC/MS and processes the resulting data (Figure 1). Results are reported via a "Data Browser" that enables the chromatographic and mass spectroscopic evidence that support each automated metabolic assignment to be rapidly reviewed locally or remotely via a secure corporate network.

It operates by comparing and contrasting each metabolised sample with a control sample—although unexpected metabolite searching may still be performed in the absence of a suitable control. Samples from *in vitro* incubations or *in vivo* dosing experiments can be quickly analysed by LC/MS, followed by a multi-dimensional data search which correlates retention time, m/z value, intensity and

EXPERIMENTAL CONDITIONS

Mass Spectrometer:	Micromass® Q-Tof micro™
Ionisation mode:	Electrospray positive ion mode
Cone voltage:	30 V
Capillary Voltage:	3 kV
Source Temperature:	120 °C
Desolvation Temperature:	270 °C
Lock mass:	Leucine enkephalin m/z 556.2771, concentration 0.5 ng/µL
Solvent Delivery System:	ACQUITY UPLC™
Column:	ACQUITY UPLC™ BEH C_{18} column, 100 x 2.1 mm id, 1.7 µm particle size
Mobile Phase A:	water + 0.1% formic acid
Mobile Phase B:	acetonitrile + 0.1% formic acid
Gradient:	0–0.25 min 100% A, 7 min 5% A, 8 min 5% A, 8.1 min 100% A
Flow rate:	400 µL/min
Injection Volume:	5 µL

Sample details: Verapamil was incubated at 10 µM using rat liver microsomes (1 mg/mL total protein concentration) for 60 minutes at 37 °C. UDPGA was also added with a concentration of 1.9 mg/mL to generate glucuronide conjugations. The reaction was then stopped by adding cold acetonitrile and left to stand at room temperature for 15 minutes. Then, it was centrifuged at 13,000 rpm for 20 minutes and the supernatant was taken for LC/MS analysis.

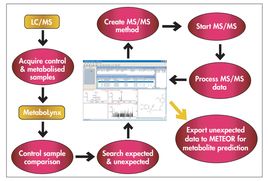

Figure 1. Describes the sequence of steps involved in Metabolynx for metabolite identification.

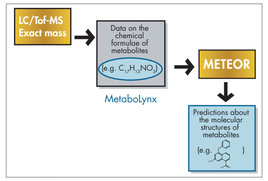

Figure 2. Workflow for metabolite analysis, detection and prediction.

components from alternative detection technologies (e.g. diode array UV or radiochemical monitoring). Comparison of analyte data with the control sample allows filtering of matrix-related peaks, which would otherwise produce an unmanageable list of false metabolite peaks.

Isotopic cluster analysis can be built into the Metabolynx™ automated processing method and is used to target potential metabolites with the desired isotope ratios. For example Cl or radiolabeled containing drugs/metabolites can be pinpointed, at low levels, within a complex matrix background to dramatically enhance specificity and increase confidence in metabolite identification.

IN SILICO METABOLITE PREDICTION, METEOR AND METABOLYNX—HOW DOES IT WORK?

METEOR predicts the likely metabolites from a parent drug structure. It uses a knowledge base of biotransformations derived by experts from literature publications and metabolism databases. The metabolites generated can be restricted via user-defined controls such as:

- Filter for metabolites that METEOR predicts to be PROBABLE via its sophisticated reasoning engine
- Filter for metabolites that have retained the radiolabel from the parent
- Filter for metabolites matching the elemental compositions generated by Metabolynx™ from LC/MS exact mass (Figure 2)

METEOR contains a direct link to the DEREK for Windows toxicity prediction software to produce an assessment of potential metabolite toxicity. It also has an integral knowledge base editor for the implementation of in-house biotransformations.

WHY EXACT MASS METABOLITE IDENTIFICATION AND DATA PROCESSING?

- Exact mass measurements enable the elemental composition of detected peaks to be confirmed for "known" drugs and their metabolites using both MS and MS/MS spectra.
- For unknowns the number of plausible elemental compositions may be restricted to a small number (or uniquely identified) with the aid of additional chemical information—e.g. the molecular formula of the parent drug and knowledge of possible metabolic pathways.

EXACT MASS FILTER WINDOW EXPLANATION

This is a very accurate and specific filter because it is based on exact mass and mass deficiencies, which are specific to each parent drug compound of interest. Each parent drug has a specific number of elements (C,H,N,O, ...). Depending on the number of each one of the elements mentioned, the drug of interest will have a very specific mass deficiency. For example if we look at a very well known pharmaceutical drug such as Verapamil, it contains the following elements; $C_{27}H_{38}N_2O_4$. This equates to a monoisotopic protonated mass of 455.2910 Da. If we take an alkyl group away (N-dealkylation, a common metabolic route) then the mass is shifted by -14.0157 Da leaving us with a monoisotopic mass of 441.2753 Da. If we now work out the delta mass difference for the 4 decimal places between Verapamil and its N-dealkylated metabolite, we are left with an exact mass deficiency of 0.2910 – 0.2753 = 0.0157 Da. Therefore, if we were to put a window of around 20 mDa we would be able to detect its N-dealkylated metabolite.

With this in mind, we can make the following hypotheses:

- All metabolites have masses within 0.25 Da of parent decimals
- They are in general within 0.1 Da if there are no major cleavages leading to much smaller fragments (as an example, the biggest single phase II biotransformation, glutathione conjugation, will lead to a mass defect difference of 0.068 Da compared to the parent drug).
- Most metabolites will fall within a 180 mDa window of the parent compound, even if certain cleavages take place in the structure to yield smaller fragments.

In Figure 3, it can be observed a series of different mass shifts, which are common from the metabolism of Verapamil.

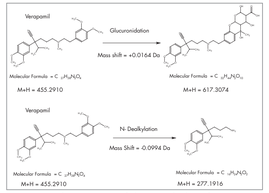

Figure 3. Shows some of the metabolites for Verapamil including cleavages, which take place giving a maximum mass deficiency of 99.4 mDa away from the parent compound. Mass deficiency shifts are very specific for each metabolite and parent drug.

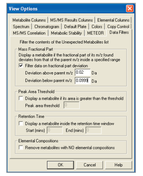

Figure 4. Shows the exact mass data filter in the Metabolynx browser.

This new development has been implemented in the Metabolynx browser as a data filter, which is carried out after the data has been processed (Figure 4). In turn, this will allow the user to set very low threshold without the worry of optimising the set up parameters to detect putative metabolites, because the unwanted peaks will be filtered at a later stage.

Important set-up processing parameters;

- Integration threshold can be defined according to the desired level of sensitivity or limit of detection (Figure 5)
- Potential metabolites intensity threshold: critical to detect minor metabolites co-eluting with major metabolite or endogenous compound.
- Exclusion Masses window: peaks observed in control will be automatically excluded in the analyte. The user will be able to review the excluded peaks in the browser. Moreover, the user may enter contaminants such as PEG manually so that if found in the analyte these peaks will then be automatically excluded.
- Elemental composition of putative metabolites is automatically calculated

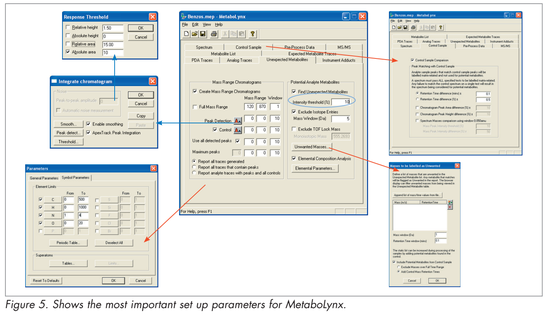

Figure 5. Shows the most important set up parameters for Metabolynx.

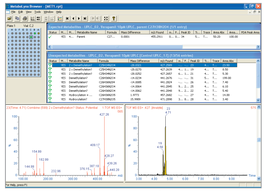

Figure 6. Metabolynx browser for Verapamil metabolites after being filtered using the exact mass filter.

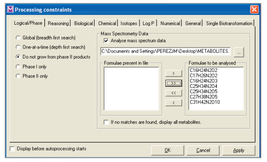

Figure 7. Exported elemental compositions from Metabolynx for further in silico metabolite prediction.

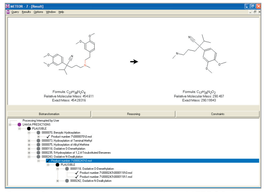

Figure 8. Metabolite prediction from METEOR for one of the Verapamil metabolites (N-Dealkylation).

RESULTS

Metabolism of Verapamil in vitro and Metabolynx/ METEOR processing

- The integration threshold was set at a spectrum intensity of 10% for metabolites of interest.
- Only the parent compound was selected as expected metabolite. Control and analyte samples were processed using mass by mass search over the defined mass range.
- 56 entries were identified in the analyte sample without any mass filtering applied.
- After applying a filter above and below the parent mass only 16 real entries were obtained which were then passed to METEOR for metabolite prediction
- All the Verapamil metabolites (Figure 6) expected from literature were identified using Metabolynx
- The elemental compositions obtained from the Metabolynx report were automatically exported to METEOR (Figure 7)
- All the metabolites detected for Verapamil were identified by METEOR following prior filtration of results generated by the in silico metabolite prediction output from METEOR
- As it can be observed in Figure 8, one of the N-Dealkylated cleavage metabolites was detected and identified from the combination of exact mass/Metabolynx/METEOR

CONCLUSIONS

- The new developments implemented for this software algorithm are a breakthrough in the field of metabolite identification.
- Data processing has been greatly improved and it is now extremely fast, even if you consider the amount of different processing steps, which take place.
- If this was done manually, it would take the user days or even weeks, to come out with such an output.
- Nowadays, rapid turn around of results, especially in discovery, is of paramount importance to be able to cope with the "bottleneck" of data processing and reporting.
- Exact mass plays an important role in metabolite identification.
- The use of exact mass and software algorithms such as the exact mass filter,which allow the user to obtain the best possible data so that decisions on whether the metabolite is real or not are not the "rate limiting step".
- The combined use of mass exclusion from control with subsequent mass filtering using 4 decimal places of the TOF provided a rapid and reliable metabolite identification process, even for minor metabolites in complex biological matrices.
- Usual problems such as missing unexpected metabolites and tedious manual comparison of analyte vs. control for MH+ MH- of every suspected metabolite are no longer an issue.
- Using METEOR will speed up the time taken to interpret the mass spectrometry data and facilitate reporting of metabolites. Moreover, it may also help to design further MS/MS experiments since we have more knowledge about the possible structure of interest.