Platform Agnostic Data Processing Routine for Targeted and Untargeted Metabolite Identification in Drug Discovery

Richard Lee,[1] Vitaly Lashin,[2] Andrey Paramonov,[2] Alexandr Sakharov,[2] Alexey Aminov[2]

[1]Advanced Chemistry Development, Inc. (ACD/Labs), 8 King Street East, Toronto, ON. M5C 1B5. Canada; [2]ACD/Labs, Moscow, Russia

INTRODUCTION

Information gained from metabolite analysis plays a critical role in early drug discovery and development. The principle method of recognizing these metabolic “hotspots” are through interpretation of mass spectrometry data, resulting in elucidation of biotransformation pathways. Although there have been a number of significant instrumental and software advancements to aid in metabolite identification, the main challenge in these studies is still structure elucidation of metabolites from the parent compounds.

Here we present, MetaSense™, a new approach for the automated identification of metabolites, which allows data from all major mass spectrometry vendors to be processed, reviewed, and uploaded to a knowledge management system. Automated file capture and processing capabilities makes it suitable for high-throughput environments with the capability for manual review and update of information by the scientist, for example modifying the identification of a metabolite, keeping the expert in control. A structure based prediction approach is used to help reduce the number of false positives. Confirmation of the site of biotransformation is checked using the available MS/MS data. The metabolite fragment mass shifts, relative to the parent MS/MS spectrum, help localize the site of biotransformation. In the cases where there is not sufficient evidence to support a single site of biotransformation, the metabolite structure can be represented using the Markush notation. All the metabolites and metadata can then be stored in a database for future use. This allows for a greater degree of collaboration between discovery and development which can save a significant amount of time and effort.

METHOD

A new ACD/Labs platform, MetaSense™, was developed for batch processing mass spectrometry data acquired from any major instrument vendor. Workflows for LC/MS^n data from several instruments, including various Orbitrap and Q-Tof analyzers, were examined. Part of the automated process involved metabolite prediction based on a regio-selectivity model, which were used as a potential metabolite target list. To complement the prediction driven approach, data driven untargeted analysis was also performed. In cases where the algorithm was not able to determine a discrete structure, the software was able to provide Markush structures. Once all data was processed, the biotransformation map and all associated mass spectra were uploaded to a knowledge management database.

The data shown is from a rat microsomal incubation. Test articles at a concentration of 10 mM in DMSO were dispensed by an acoustic dispenser (25 nL) to 25 µL 10 mM phosphate buffer (pH 7.4) containing 2 mM NADPH. This solution (12.5 µL) was added to 12.5 µL rat liver microsomal protein (1 mg/mL). At specific time points (0, 2, 5, and 10 min), the reactions were terminated by the addition of 10 µL acetonitrile/formic acid (93:7). The samples were analyzed on an Elite Hybrid Velos Pro Ion Trap/Orbitrap (ThermoFisher Scientific, CA, USA) mass spectrometer equipped with an electrospray (ESI) source operating in positive mode electrospray ionization. Data-dependent acquisition based on a list of m/z values of potential metabolites was applied. The resolution was set to 30,000 in full scan mode and 15,000 for high energy collision dissociation (HCD) MS2.

AUTOMATED WORKFLOW

Figure 1. Workflow diagram outlining a simplified data processing approach for metabolite detection and identification on the MetaSense platform—automation is available at several stages as indicated in the scheme.

A set of high resolution LC/MS^n data files representing a study across multiple incubation time points were processed post-acquisition. Data files, along with parent structure files, were automatically processed within the new software routine. Structures of possible phase 1 and phase 2 metabolites were generated from an assembly based metabolism model. Components were detected based on this list. To complement the prediction driven approach, a non-targeted unexpected metabolite extraction procedure was integrated into the overall processing routine, employing a mass difference filter within a component detection algorithm. Metabolites were initially identified based on their accurate mass and theoretical isotopic distribution calculated from molecular formulae. Their subsequent MS2 and MS3 spectra were extracted, where available, and used later to verify metabolite structures. The software was able to assign fragment ions of the parent and metabolites to their respective MS2 spectra. Structures of metabolites were verified and scores were provided by comparing the assigned fragments and common neutral losses between it and the parent. In situations where the software was not able to provide a discrete structure, Markush notations were employed until further information was available which allowed for manual changes to the substructure.

Once the data was processed, the software routine combined both predicted and unexpected metabolites into a single biotransformation map, where all mass spectra were associated to each of the structures, and uploaded to a knowledge management system for data review. Peak areas of the parent and metabolites were monitored across all incubation time points to generate a summary table of all detected components. Peak areas from the summary table were graphically represented as a bio-kinetic (or stability) plot.

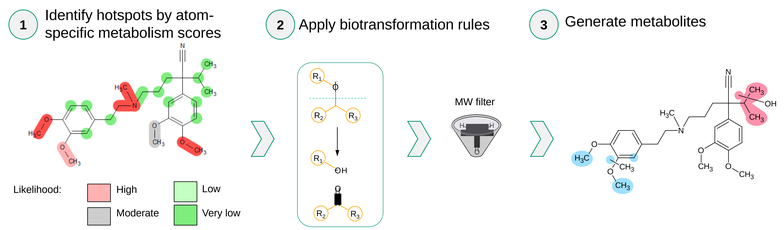

Figure 2. The three main stages of the metabolism prediction for Verapamil as predicted by ACD/Percepta™.

METABOLISM MODEL

A structure based prediction approach was used to target expected metabolites. Metabolite prediction is performed using the ‘regioselectivity of metabolism’ algorithm implemented in ACD/Percepta™ software. This algorithm consists of several steps:

In the first step, a probabilistic statistical model is applied to estimate the likelihood of a metabolic reaction taking place at each potential site of metabolism in the compound of interest. Results are reported as a metabolism score ranging from 0 to 1 that takes into account both calculated probability of metabolic reaction and Reliability Index (RI)—a quantitative measure of prediction confidence. The figure above illustrates these predictions for Verapamil, color-coded by score values for easy visualization.

Once potential metabolic hotspots are identified, they are checked against a database of biotransformation rules to identify what types of metabolic reactions are defined for the respective site of metabolism, taking into account its chemical neighborhood. Finally, selected biotransformation rules are applied to generate metabolite structures.

Depending on user-specified settings, the algorithm may run multiple stages of predictions, submitting the metabolites obtained in the first stage for further calculations to obtain the full metabolic tree, which may include Phase I and Phase II metabolites.

DISCUSSION

A series of datasets representing incubation studies comprised of 4 time intervals were batch processed using the described software routine, including verapamil. Once the data was processed, the complete interpreted results were uploaded to the knowledge management database (Spectrus DB) as shown in Figure 3.

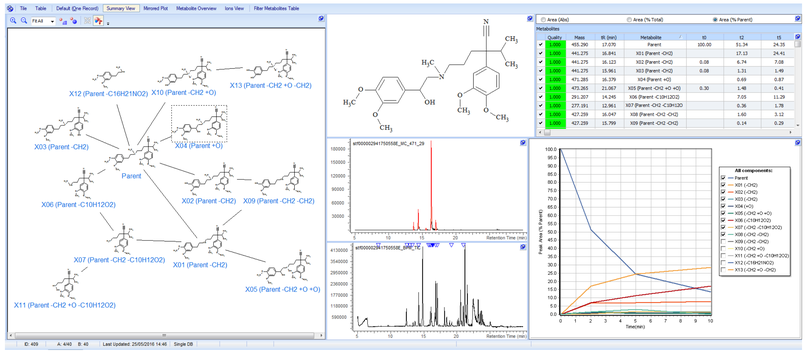

Figure 3. Summary review interface featuring biotransformation map, corresponding chromatograms, summary table of metabolites, and stability and kinetic plot of parent and metabolites, respectively.

The predicted metabolites were restricted to human specific phase 1 and 2. Processing parameters were fixed to report the top 8 metabolites based on the peak area from their respective extracted ion chromatograms. Structures of the metabolites verified by MSMS interpretation were tabulated in the summary table where the corresponding stability (parent) and the biokinetic plot (metabolites) were generated. Prior to comparing the MS2 spectra of the parent and metabolite, fragmentation analysis was performed on the parent structure, to assign fragments to the MS2 spectrum. Once assigned, the parent and metabolite MS2 spectra were compared, where common fragments and m/z pairs between the parent and metabolite were assessed. Pairs of ions were grouped together by correlating common mass shifts, i.e., if two ions were separated by Δ14 (demethylation), as shown in Figure 4.

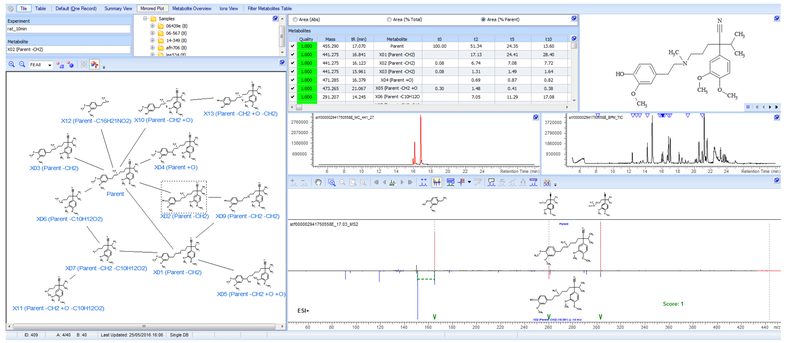

Figure 4. Metabolite centric view of the processed results presenting mirrored display between the MS2 spectra of the fragment assigned parent and metabolite.

DATA VISUALIZATION

A summary of the data was automatically generated and updated to Spectrus DB upon completion of the data processing routine. To communicate the results and increase collaborative efforts, a new java script based web interface was designed to view the results as shown in Figure 5. This new web enabled interface is browser-independent. Each of the elements within the browser have been developed as an independent “widget” which can be incorporated as into a 3rd party web interface. The viewer in this case includes a display of the biotransformation map as the main feature. Users can perform metadata search, and more importantly, structures can be searched by substructure, similar structure, and exact structure search via drawing applet within the web interface.

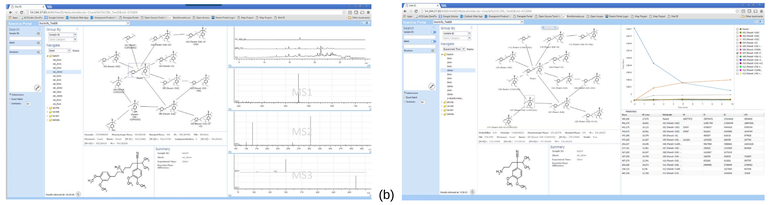

Figure 5. Web interface displaying data retrieved from Spectrus DB. a) Chromatographic and spectral data view. b) Stability and kinetic plot of parent and metabolites tracked across an incubation study, and associated summary of peak areas.

CONCLUSION

MetaSense™ provides an automated and efficient platform for metabolism studies carried out on data from different instrument vendors, without compromising the quality of interpretation. The process of identifying potential metabolite peaks, interpretation of fragments, and generation of biotransformation maps significantly improves the speed of data interpretation and thereby the quantity of samples that can be analyzed. Metabolite ID and DMPK experts responsible for metabolism studies would save significant time and effort by applying this software aided approach.

Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com