Rapid and Confident Metabolite Profiling and Identification using Bench-Top Orbitrap Q Exactive and Compound Discoverer

Kate Comstock,1 Caroline Ding,1 Tim Stratton,1 Kelly Wang,2 Gene Eiserberg2
1Thermo Fisher Scientific, San Jose, CA; 2Gilead Sciences, Foster City, CA
Overview

Purpose: Single software solution for HRAM LC/MS data processing for confident and rapid metabolite identification and structure elucidation.

Methods: High-resolution LC/MS and novel software for metabolite profiling and structure identification.

Results: Compound Discoverer enables fast, efficient, and confident metabolite profiling in an all-in-one UHPLC/HR-MS/MS platform.

Introduction

In vitro drug metabolite identification by LC coupled with HRAM MS is an essential component of drug discovery to select and prioritize compounds with the best chance of success. Data processing and data collation are the current rate-limiting steps of the MetID process.

This study demonstrates how using Thermo Scientific™ Compound Discoverer™ Software improves the MetID process. Compound Discoverer software is pipeline-based where each node performs a discrete function. The user has the flexibility to arrange appropriate nodes into a workflow that produces the desired information. A branched pipeline can be used to process a data set in multiple ways to extract maximal information and to bring all the data together for review and reporting.

Methods

Sample Preparation

Darunavir samples (50 uM) were incubated in human liver microsomes (1 mg/mL) fortified with NADPH (1mM) for 0 and 2 hours.

Liquid Chromatography

Thermo Scientific™ Ultimate™ 3000 RS UHPLC system
Column: Thermo Scientific™ Accucore™ C18 (100 x 2.1 mm), 2.6 μm
Column Temperature: 35 °C
Gradient Solvents: A: H₂O/0.1% Formic Acid; B: ACN/0.1% Formic Acid
Flow rate: 500 μl/min
Injection Volume: 5 μl

<table>
<thead>
<tr>
<th>LC gradient:</th>
<th>0.0</th>
<th>14.0</th>
<th>18.0</th>
<th>20.0</th>
<th>22.0</th>
<th>22.1</th>
<th>26.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>A%</td>
<td>95</td>
<td>60</td>
<td>60</td>
<td>20</td>
<td>20</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>B%</td>
<td>5</td>
<td>40</td>
<td>40</td>
<td>80</td>
<td>80</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Mass Spectrometry

The MS analyses were carried out on a Thermo Scientific Q Exactive™ mass spectrometer using the electrospray technique in the positive ion mode. High-resolution accurate mass (HRAM) full-scan MS and top 3 MS/MS spectra were collected in a data-dependent fashion at a resolving power of 70,000 and 17,500 at FWHM m/z 200, respectively. The Stepped NCE (Normalized Collision Energy) setting was 15, 20, 25.

Data Analysis

A study was created in Compound Discoverer for Darunavir. The study files include raw files from t₀h, t₂h with and without NADPH and a Darunavir standard injection. The standard injection was used to ensure there were no major impurities in the Darunavir drug. Compound Discoverer software employs a drag-and-drop user interface to facilitate quick and flexible file relationship assignments for downstream control comparison and retention time alignment (Figure 1).
Rapid and Confident Metabolite Profiling and Identification using Bench-Top Orbitrap Q Exactive and Compound Discoverer

Gradient Solvents: A: H2O/0.1% Formic Acid; B: ACN/0.1% Formic Acid

Injection was used to ensure there were no major impurities in the Darunavir drug.

Compound Discoverer software employs a drag-and-drop user interface to facilitate quick and flexible file relationship assignments for downstream control comparison and retention time alignment.

Results

Methods

Darunavir samples (50 uM) were incubated in human liver microsomes (1 mg/mL) fortified with NADPH (1mM) for 0 and 2 hours.

Control_t0h, NADPH_t0h, and NADPH_t2h for RT alignment and comparison.

The three files, Control_t0h, NADPH_t0h, and NADPH_t2h, were processed in one analysis using Compound Discoverer software. The node-based processing workflow (Figure 2) included both expected metabolite detection and mass-defect-filtered unknown component detection. The Mass Defect Filter node is a spectra filtering tool and was used before the Unknown Detector node. The mass defect filters were created on the basis of parent de-alkylation predictions and expected transformations with a mass tolerance of +/- 50 Da and a mass defect tolerance of +/- 50 mmu. The Expected Finder node includes de-alkylation and de-arylation predictions based on the structure of Darunavir. Common phase I bio-transformations were automatically combined on the basis of the number of maximum occurrences for each transformation and the maximum number of combinatorial steps. The Fragment Ion Search (FISh) Scoring node searches MS/MS spectra for fragment ions in common with the parent fragment ions or those that are shifted by a predicted biotransformation. A positive match confirms the presence of a compound that is structurally related to the parent and provides the location of the modification on the basis of the shifted fragments. Since Darunavir contains one sulfur (S) atom and the full-scan MS data was run at a resolving power of 70,000, the Isotope Ratio Tracer node was used in the workflow to create a 1S fine isotope trace as an additional trace for parent-related compound detection, and UV traces were also included as orthogonal traces. The Compare with Control feature was enabled in the workflow to compare compounds in the three files.

FIGURE 2. Node based processing workflow in Compound Discoverer Software

Results

Compound Discoverer software uses both metabolite prediction and advanced component detection to discover both predicted and unexpected metabolites in the data. Advanced, automated tools that use isotope ratios, isotope fine structure, and correlation of fragment peaks provide confidence in the identification of drug related components and site of modification. Results are consolidated into one peak table. Furthermore, the same peaks found in the sample and control files are collated and consolidated. Overlay chromatograms of the same compound found in NADPH_t2h, NADPH_t0h, and Control_t0h are displayed in the chromatogram window for visual inspection (Figure 3). Adducts are automatically detected and grouped for each compound (Figure 4). Adduct grouping greatly reduces false positives from compound hits.
Rapid and Confident Metabolite Profiling and Identification using Bench-Top Orbitrap Q Exactive and Compound Discoverer

Mass Spectrometry

LC gradient:

Column Temperature: 35 °C

Gradient Solvents:  A: H2O/0.1% Formic Acid;

Column:  Thermo Scientific™ Accucore™ C18 (100 x 2.1 mm), 2.6 µm

Flow rate: 500 µl/min

injection was used to ensure there were no major impurities in the Darunavir drug.

This study demonstrates how using Thermo Scientific™ Compound Discoverer™ Software component of drug discovery to select and prioritize compounds with the best chance of

In vitro drug metabolite identification by LC coupled with HRAM MS is an essential

identification.

Overview

together for review and reporting.

Data Analysis

at a resolving power of 70,000 and 17,500 at FWHM

time alignment (Figure 1).

Liquid Chromatography

95

60

20

20

m/z

200, respectively. The Stepped

Darunavir

Control_t0h are displayed in the chromatogram window for visual inspection (Figure 3).

Modification. Results are consolidated into one peak table. Furthermore, the same peaks

Compound Discoverer software uses both metabolite prediction and advanced component

Results

FIGURE 2. Node based processing workflow in Compound Discoverer Software

analysis using Compound Discoverer software. The node-based processing workflow

FIGURE 1. Easy grouping of files by dragging and dropping the files to the

FIGURE 3. Overlaid XIC provides visualization of how the same compound
changes in three files.

FIGURE 4. Automatic adduct grouping reduces false positive assignments. Isotope patterns are automatically considered with color coding to represent the fidelity of isotope pattern fit for assigned elemental composition.

For each expected metabolite, FISh Scoring automatically compares the MS/MS fragmentation spectra to that of the parent compound and annotates the spectra with matching fragment structures (color-coded in green) and biotransformation shifted fragments (color-coded in blue). This provides a quick visual indication of which part of the molecule has been modified and which part remains unchanged for a rapid determination of the modification location (Figure 5).

FIGURE 5. MSMS fragments are compared to the parent and structures automatically annotated to help determine the location of the modification.

TABLE 1. Darunavir in vitro metabolites (area threshold above 1% of parent in

<table>
<thead>
<tr>
<th>RT (min)</th>
<th>M5</th>
<th>M3</th>
<th>M1</th>
<th>12.66</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>408.19473</td>
<td>548.24195</td>
<td>563.23014</td>
<td>579.22505</td>
<td>391.19296</td>
<td>579.22505</td>
</tr>
<tr>
<td>C27H37N3O8S</td>
<td>C27H37N3O9S</td>
<td>C27H37N3O9S</td>
<td>C20H29N3O4S</td>
<td>-</td>
<td>+O2</td>
</tr>
<tr>
<td>13099948</td>
<td>43961061</td>
<td>3192335</td>
<td>58349</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NADPH_t0h</td>
<td>91409870</td>
<td>10302</td>
<td>16909</td>
<td>16074</td>
<td>0</td>
</tr>
<tr>
<td>Control_t0h</td>
<td>69281472</td>
<td>10302</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

and MS/MS data from the bench top Q Exactive instrument, and the advanced processing annotations algorithm for the proposed structure in Compound Discoverer. In the example

productivity.

• Untargeted components increases the confidence of compound identification.

an isotope ratio tracer, a targeted search for parent compounds and their transformation interest is missed.

Metabolite ID is a complex and time-consuming process. By combining the HRAM full-scan MS

Conclusion

FIGURE 8. FISh re-annotated MS/MS spectrum based on proposed putative structure

Putative metabolite structure assignment was confirmed by re-running the FISh matching and

detected sample components.

Flexible data review and multiple monitor support provides simultaneous viewing of

detected sample components.
Compound Discoverer software supports multiple monitors and includes a data review window that is easily customized for the MetID review process. The Consolidated Peaks table shows consolidated peaks detected from the Expected Finder and/or Unknown Detector nodes and from all three files in the same table simultaneously. The tables in Compound Discoverer support multiple-column sorting and multiple-property filtering. An area threshold filter of 1% of the parent compound in NADPH_t0h and a RT range filter of 2-17min were applied to reduce the number of peaks for data review. Furthermore, the In Control status was used as a filtering condition to filter out peaks that were found in both of the 10h samples (NADPH_t0h and Control_t10h) and that did not change significantly in the 12h sample. The following ten putative metabolites were identified in the NADPH_t12h sample in addition to the parent compound (Table 1).

**TABLE 1. Darunavir in vitro metabolites (area threshold above 1% of parent in NADPH_t10h) detected by HRAM and Compound Discoverer.**

<table>
<thead>
<tr>
<th>Name</th>
<th>RT [min]</th>
<th>Apex m/z</th>
<th>MW</th>
<th>Formula</th>
<th>Formula change</th>
<th>Comment</th>
<th>NADPH_t12h</th>
<th>NADPH_t10h</th>
<th>Control_t10h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent</td>
<td>14.46</td>
<td>548.24195</td>
<td>547.2352</td>
<td>C27H37N3O8S</td>
<td></td>
<td>Darunavir</td>
<td>43861061</td>
<td>91409870</td>
<td>69281472</td>
</tr>
<tr>
<td>M1</td>
<td>8.08</td>
<td>408.19473</td>
<td>407.18788</td>
<td>C20H29N3O4S</td>
<td>-(C7H8O4)O2 oxidation + oxidation</td>
<td>3192335</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>9.74</td>
<td>408.19501</td>
<td>407.18788</td>
<td>C20H29N3O4S</td>
<td>-(C7H8O3)O2 oxidation</td>
<td>1604308</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>10.48</td>
<td>580.23216</td>
<td>579.22505</td>
<td>C27H37N3O8S</td>
<td>O2 oxidation + oxidation</td>
<td>899572</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>10.55</td>
<td>392.19975</td>
<td>391.19296</td>
<td>C20H29N3O4S</td>
<td>-(C7H8O4)O2 oxidation</td>
<td>13789370</td>
<td>58349</td>
<td>12381</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>10.85</td>
<td>580.23224</td>
<td>579.22505</td>
<td>C27H37N3O8S</td>
<td>O2 oxidation + oxidation</td>
<td>952987</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td>11.74</td>
<td>564.23718</td>
<td>563.23014</td>
<td>C27H37N3O8S</td>
<td>O oxidation</td>
<td>13099948</td>
<td>33394</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>12.66</td>
<td>580.23218</td>
<td>579.22505</td>
<td>C27H37N3O8S</td>
<td>O2 oxidation + oxidation</td>
<td>1268479</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M8</td>
<td>13.10</td>
<td>564.23712</td>
<td>563.23014</td>
<td>C27H37N3O8S</td>
<td>O oxidation</td>
<td>3934473</td>
<td>16074</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M9</td>
<td>13.32</td>
<td>564.23.724</td>
<td>563.23014</td>
<td>C27H37N3O8S</td>
<td>O oxidation</td>
<td>3884732</td>
<td>16909</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>M10</td>
<td>13.62</td>
<td>564.23712</td>
<td>563.23014</td>
<td>C27H37N3O8S</td>
<td>O oxidation</td>
<td>9237332</td>
<td>51973</td>
<td>10302</td>
<td></td>
</tr>
</tbody>
</table>

UV traces were used in addition to the various filtered mass chromatograms to help correlate back to non-LC/MS-based studies. Figure 6 shows overlaid UV traces of NADPH_t0h and NADPH_t12h that are added and stacked on top of the combined MS trace of the 10 putative metabolites and parent.

**FIGURE 6. Stacked view of combined ms trace from the 10 major metabolites and overlay of UV traces from NADPH_t0h and NADPH_t12h.**

Metabolite structure assignments were facilitated by FISh annotations. An example is given below (Figure 7) showing how a comparison of the FISh-annotated MS/MS spectrum of the oxidated Darunavir metabolite to the parent HCD MS2 spectrum was used to help determine the site of transformation.
Methods

success. Data processing and data collation are the current rate-limiting steps of the MetID component of drug discovery to select and prioritize compounds with the best chance of In vitro drug metabolite identification by LC coupled with HRAM MS is an essential rapid metabolite identification and structure elucidation.

Overview

node performs a discrete function. The user has the flexibility to arrange appropriate nodes improving the MetID process. Compound Discoverer software is pipeline-based where each NCE (Normalized Collision Energy) setting was 15, 20, 25. at a resolving power of 70,000 and 17,500 at FWHM (HRAM) full-scan MS and top 3 MS/MS spectra were collected in a data-dependent fashion time alignment (Figure 1).

Thermo ScientificTM

MW 547.23522
Darunavir

Results are consolidated into one peak table. Furthermore, the same peaks grouping greatly reduces false positives from compound hits. Advanced, chromatic modification on the basis of the shifted fragments. Since Darunavir contains one sulfur modification. The ability to extract and support orthogonal pieces of data including analog trace data (from UV, PDA, or other analog detectors including radioisotope detectors) ensures no compound of interest is missed. Resolution aware and accurate-mass empowered algorithms that include mass defect filtering, an isotope ratio tracer, a targeted search for parent compounds and their transformation products based on expected modifications and de-alkylation predictions, and the ability to detect untargeted components increases the confidence of compound identification. Easy to use software interface for method development and sample comparison increases productivity. Flexible data review and multiple monitor support provides simultaneous viewing of chromatographic traces, mass spectra, and data result tables.

Conclusion

Metabolite ID is a complex and time-consuming process. By combining the HRAM full-scan MS and MS/MS data from the bench top Q Exactive instrument, and the advanced processing algorithms in the Compound Discoverer software, a previously tedious process that could take an experienced MetID scientist a few days can be reduced to one day.

• A node-based processing workflow provides flexibility in data processing.
• The ability to extract and support orthogonal pieces of data including analog trace data (from UV, PDA, or other analog detectors including radioisotope detectors) ensures no compound of interest is missed.
• Resolution aware and accurate-mass empowered algorithms that include mass defect filtering, an isotope ratio tracer, a targeted search for parent compounds and their transformation products based on expected modifications and de-alkylation predictions, and the ability to detect untargeted components increases the confidence of compound identification.

• Easy to use software interface for method development and sample comparison increases productivity.
• Flexible data review and multiple monitor support provides simultaneous viewing of chromatographic traces, mass spectra, and data result tables.
• Fully integrated fragmentation prediction from Mass Frontier™ and automatic fragment ion matching with annotations help elucidate the localization of transformations. This leads to more confident structure determinations of the expected transformation products and unknown detected sample components.