Smart Note

The ultimate toolset for transforming small-molecule MS data into meaningful insights

My small-molecule MS datasets are incredibly complex, making it challenging to determine what's in my samples, in what amounts, and most importantly, what it indicates about the system I'm studying. Is there a solution that addresses these challenges?

Today's high resolution, accurate mass (HRAM) mass spectrometers (MS) provide extensive, rich, raw mass-spectral data that require sophisticated toolsets to turn them into useful insights about the system under investigation. With the ability to make retention time corrections, predict elemental formulas, detect features, eliminate redundant data, annotate unknowns, find relevant differences, and place information into its biological or process context (Figure 1), Thermo Scientific™ Compound Discoverer™ software is the most complete solution for processing untargeted small-molecule data. The software includes powerful tools for:

- Unknown peak detection
- Advanced statistical analysis
- Interactive data visualization
- Compound annotation
- Integrated database and spectral library searching
- Biochemical pathway mapping
- Untargeted stable isotope labeling analyses

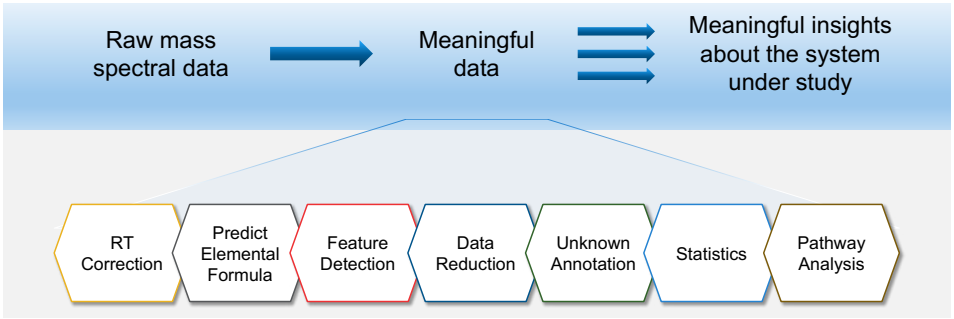

Figure 1. Transforming complex and feature-rich mass spectral data into insights requires a powerful toolset which not only aids in simplifying the ability to take raw data and convert this into actionable knowledge, but also simplifies this process and makes it available to all users within a laboratory.

ThermoFisher SCIENTIFIC

How does Compound Discoverer software simplify setup of small-molecule analyses?

Compound Discoverer software includes all the features needed to get users started quickly, as well as the flexibility to allow customization to address specific needs. Four integrated modules work together to provide a start-to-finish workflow for collecting and processing small-molecule data: Study Manager, Workflow Editor and Processing Engine, Results Analysis, and Annotation and Biological Interpretation. Further, using the Study Wizard, analyses can be setup in four easy steps: 1) import the sample list, 2) select the workflow template, 3) define the sample types, and 4) define the study variables.

The software includes numerous predefined workflow templates covering most small-molecule research applications:

- Metabolomics
- Stable isotope labeling
- Drug metabolism
- Environmental and food research
- Forensics
- Natural products
- Impurity and degradant identification

Figure 2 shows some of the metabolomics templates included with the software. Any of the templates can be modified to meet specific experimental goals.

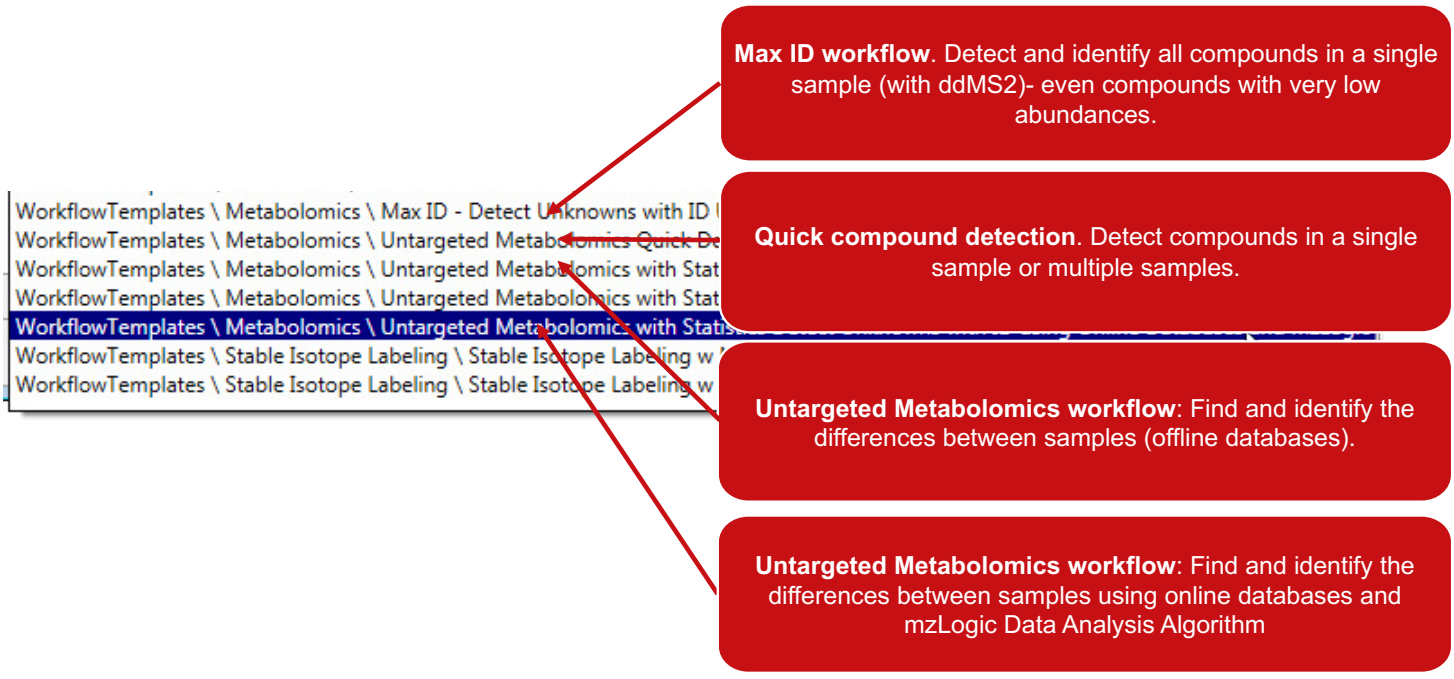

Figure 2. Examples of workflow templates for metabolomics applications.

For laboratories wanting to tailor workflows to their specific applications and compounds, the Workflow Editor is an easy-to-use tool for building custom workflows using node and pipeline architecture (Figure 3). Nodes that perform specific functions can be selected and chained together using “drag and drop” actions. In addition, workflow connections can specify intelligent data processing in branched workflows. Though the software includes numerous nodes, users can create their own, and additional third-party nodes are accessible through the software. If the user has programming skills and wants a custom node, the software provides node-scripting capabilities.

The Study Manager simplifies setup of sample relationships before MS analysis, ensuring that navigating and interrogating acquired data is easier. Study Manager provides for biological and technical replicates, nested study designs, and custom study factors. Figure 4 shows an example of a study of two yeast strains at seven time points. By clicking on check boxes, the user can quickly review processed data to visualize how it’s related to the study factors specified.

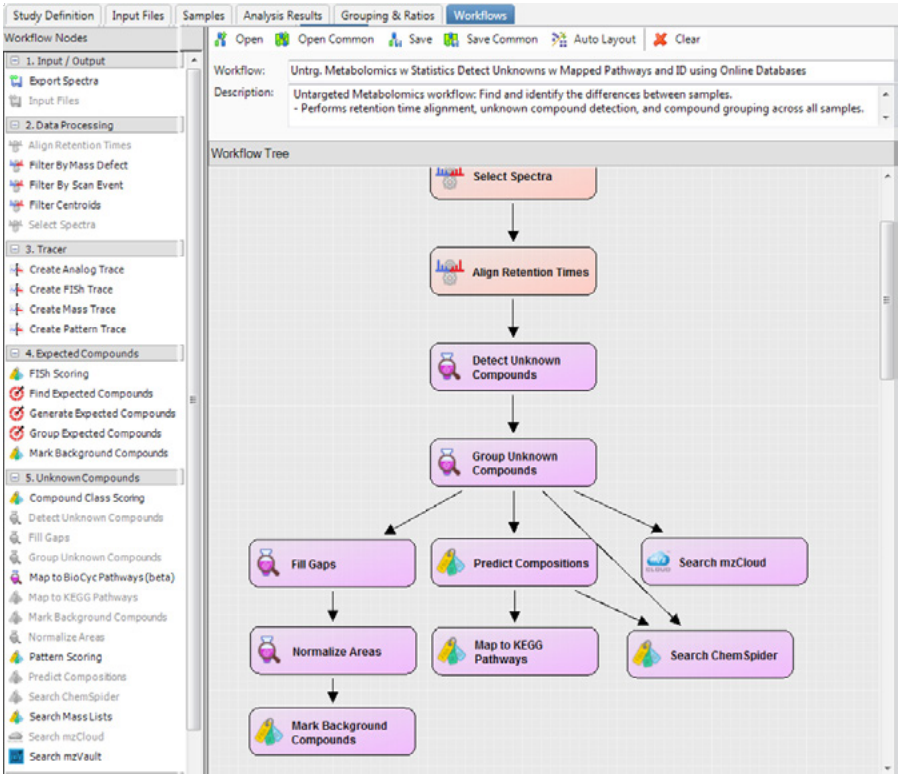

Figure 3. The workflow editor makes it easy to set up and visualize data processing workflows. Nodes on the left side of the window can be dragged and dropped into the workflow tree, allowing data to be processed as required by any given study.

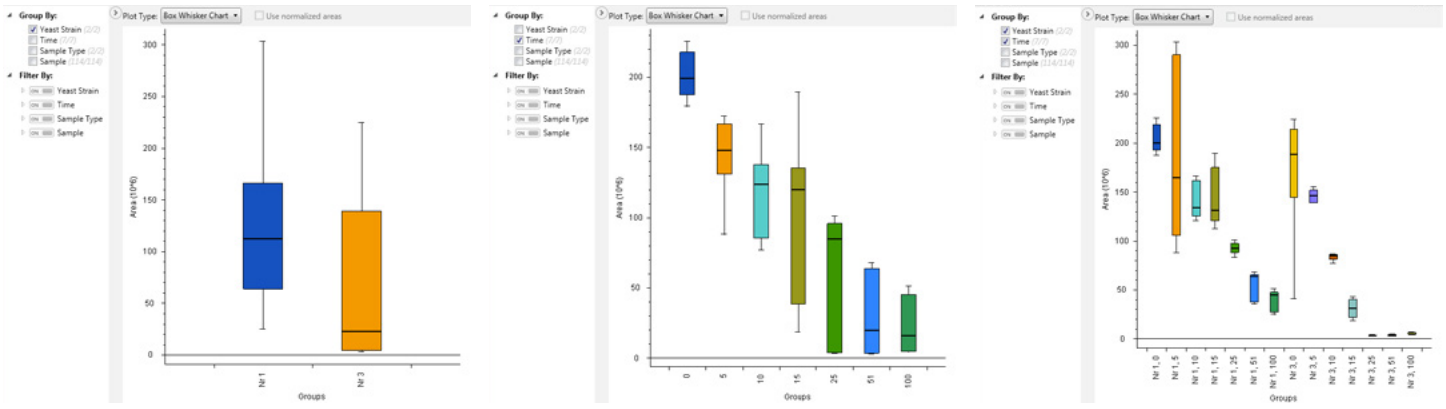

Figure 4. The Study Manager allows the user to set up relationships between samples before MS analysis, streamlining subsequent data navigation and review using checkboxes. Shown are bar charts of results by strain and time, sample by time, and both strain and sample by time.

How can I reduce the amount and complexity of MS data acquired to streamline data processing and improve quality of results?

Much of the MS data collected during small-molecule experiments is irrelevant because most of it—as much as 80%—are signals from background, not the compounds of interest. In addition, multiple ions, for example adduct ions, are often observed for a single compound, leading to redundant data and difficulties in subsequent data interpretation.

By removing background and grouping redundancy, Compound Discoverer software reduces the millions of raw data points in a file to hundreds of thousands of features and thousands of compounds (Figure 5). This rigorous data reduction step makes it substantially easier to process and interpret data.

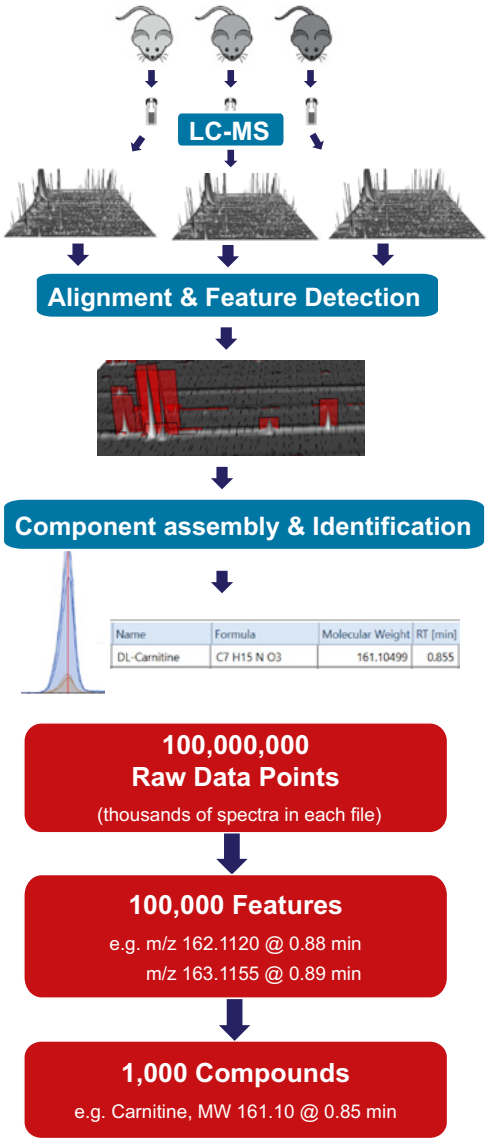

Figure 5. Data reduction: from raw data to compounds.

Thermo Scientific™ AcquireX intelligent data acquisition for Thermo Scientific™ Orbitrap ID-X™ Tribrid™ mass spectrometers goes a step further by streamlining collection of MS/MS data, avoiding the need to process hundreds of thousands of features to uncover the components of interest. To avoid triggering MS/MS data collection on background and redundant components, AcquireX software develops and deploys automatically generated inclusion- and exclusion lists as repetitive samples are injected (Figure 6).

As presented in Figure 6, the software begins by determining what is background and what is of interest by comparing the full scan data obtained for the blank and the study sample. In addition, the full scan sample analysis determines the precursor ions to include in the following MS/MS scan, avoiding collection of MS/MS data for related adducts. With each repetitive sample injection, the software automatically updates the inclusion and exclusion lists, preventing the instrument from triggering on components for which it has already collected data, and instead triggering collection of MS/MS data for lower-intensity peaks.

As shown in Figure 7, AcquireX data acquisition increases mass spectrometer efficiency, producing significantly more high-quality MS/MS information for lower abundance components. The improvement in information content leads to increased compound detection and identification (Figure 8).

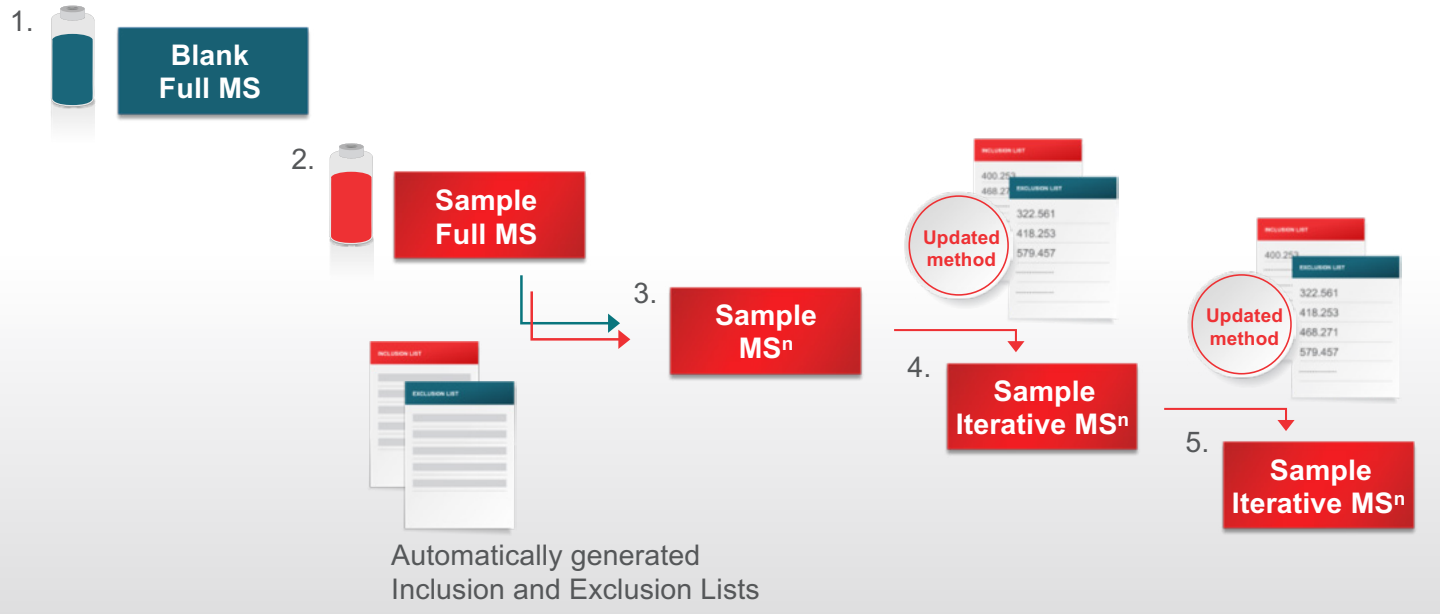

Figure 6. AcquireX data acquisition enables the user to dig deeper and deeper into their sample with each injection.

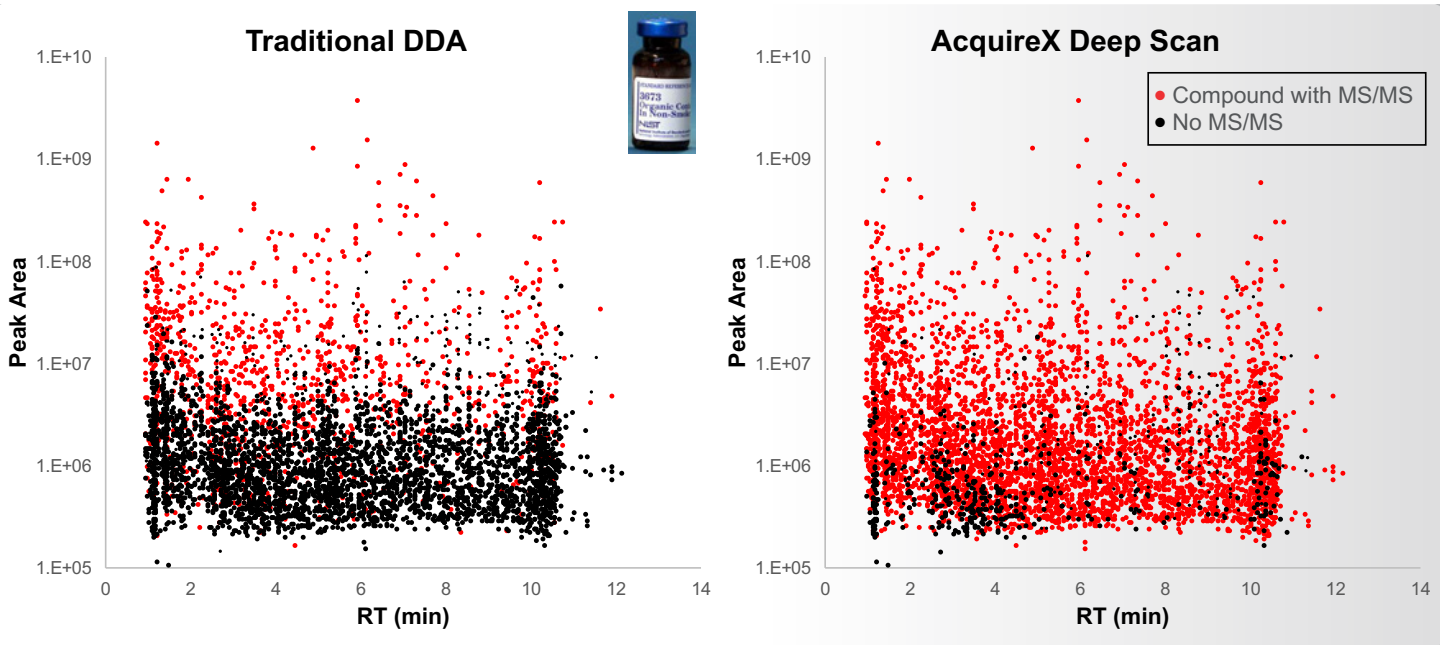

Figure 7. Comparison of traditional data-dependent acquisition (DDA) and AcquireX data acquisition of a non-smoker's urine shows massively increased depth of MS/MS fragmentation coverage for the AcquireX approach after only four injections. The red dots show components for which MS/MS scans were collected.

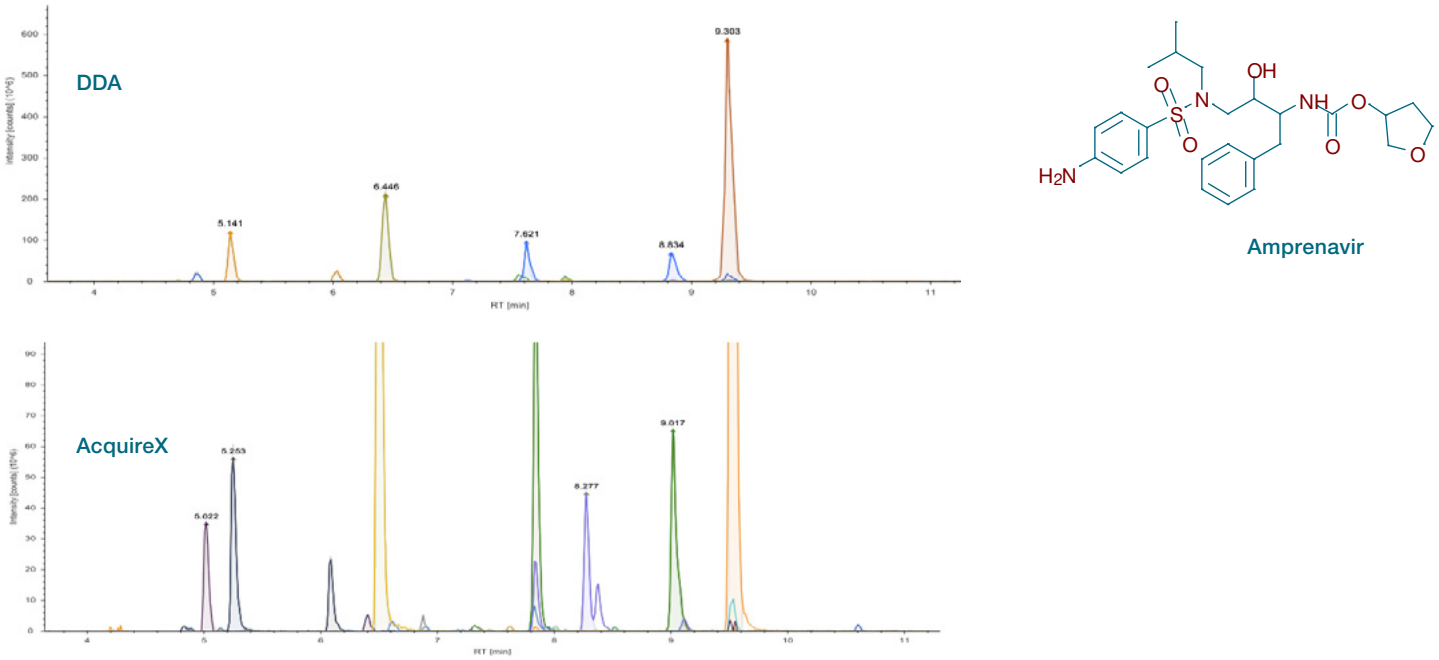

Figure 8. The signal for the drug Amprenavir spiked at 0.1 µM in bile and its metabolite is very small compared to the bile signal itself. When DDA was used (top) only eight metabolites were detected, compared to all 21 known, using AcquireX data acquisition (bottom).

What tools can help me explore and visualize my complex MS datasets?

The Results Analysis module provides sophisticated capabilities for reviewing and annotating data, and for performing statistical analyses in powerful context-sensitive displays. Shown in Figure 9, the results table view of data enables the user to obtain more information for each compound selected in the table.

Annotation results are displayed as color-coded rectangles that flag the success of the identification approaches used, for example a Thermo Scientific™ mzCloud™ mass spectral database search and chemical formula match, in the order specified by the user:

- Green represents a high-confidence identification
- Gray indicates there is not enough information to make an identification
- Orange indicates that there's more than one answer that matches the data
- Red flags a problem that may render the identification unusable

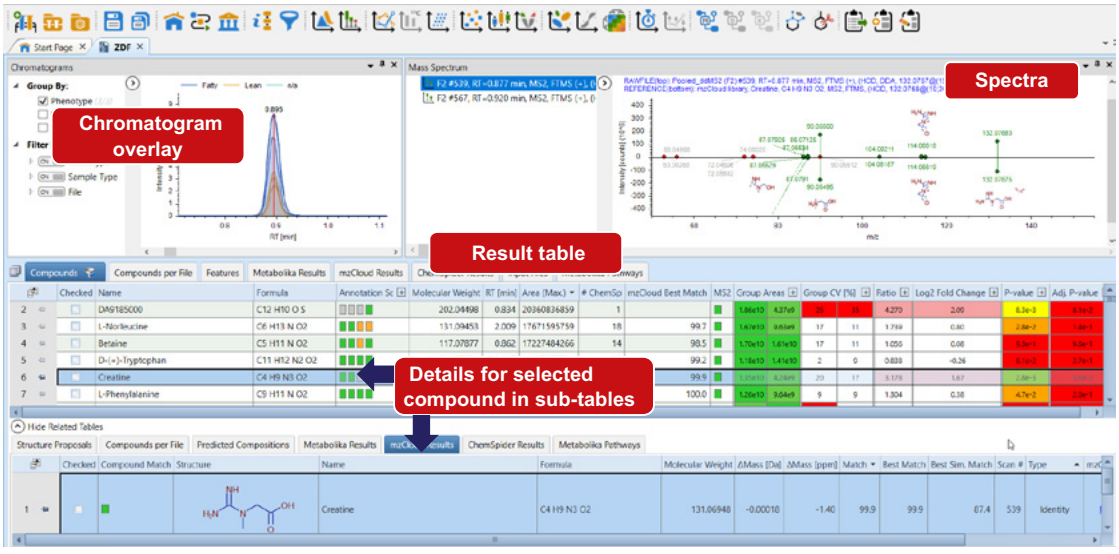

Figure 9. Results table review of an mzCloud library search showing the structure of the compound identified as creatine (bottom), and a comparison of the library spectrum to the sample spectrum (upper right). Other useful information such as formula and molecular weight are displayed.

Statistical tools enable the user to review their data in various plots to assess if the differences observed between samples are statistically significant. As components are highlighted in the results table or a statistical plot, the user can toggle to other views, allowing rapid visualization of interesting components in many ways (Figure 10). Reviewed components are checked, allowing the user to track their review, and a link back to the raw data permits results verification. Though Compound Discoverer software provides for relative quantitation, the information for any checked compound, such as retention time, mass and MS/MS information, can be directly exported to Thermo Scientific™ TraceFinder™ software for targeted screening and/or quantitative analysis.

What tools are provided to help me annotate small-molecule data?

From formula, fine Isotope, and MS^2-based annotation, to spectral library and retention time database searches, and ranked structural-similarity matches, Compound Discoverer software provides a rich toolset for making high-confidence annotations.

Thermo Scientific™ Orbitrap™ mass spectrometers produce high-resolution data that resolves elemental composition, including fine isotopes, permitting high-confidence elemental formula calculation. The software can also access MS/MS spectral libraries, including mzCloud, Thermo Scientific™ mzVault™, and import and use publicly available spectral libraries in the NIST MSP format, to compare experimentally acquired fragments with those of standards. Searches of ChemSpider or a local database can be used to make matches based on formula or accurate mass. When none of these searches provide a high-confidence match, the software includes algorithms for scoring structural candidates, including Fragment Ion Search (FISh) scoring and the Thermo Scientific™ mzLogic™ algorithm.

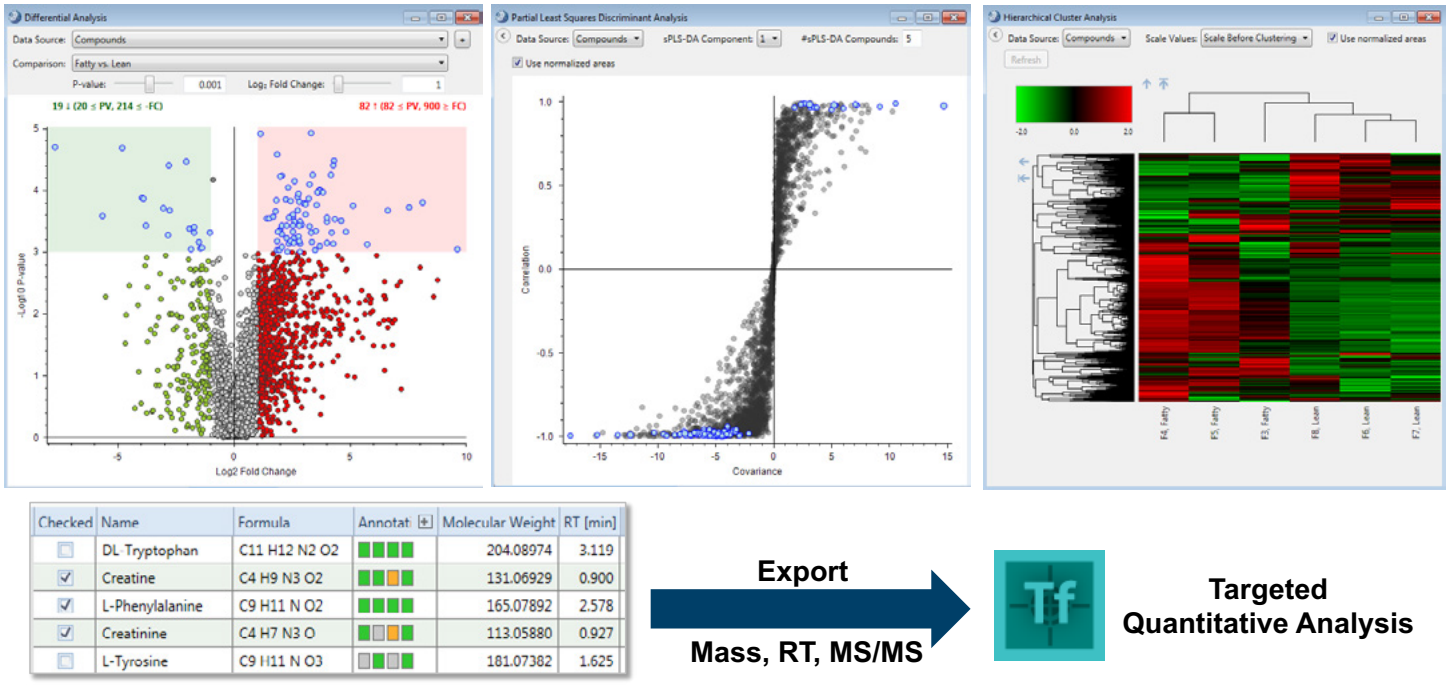

Figure 10. Interactive visualization of results allows toggling between statistical views. Results can be exported to TraceFinder software for targeted quantitation.

How is the mzCloud mass spectral database different than other spectral libraries?

The mzCloud mass spectral database is an extensively curated, high-quality mass-spectral fragmentation database that contains numerous compounds (currently over 17,600) with extensive metadata for each entry. Entries include exhaustive high-resolution MS/MS and multi-stage MS^n high-resolution accurate-mass (HRAM) spectra (currently over 6.3 million) that have been acquired at various collision energies and using different fragmentation techniques (Figure 11). Each raw mass spectrum has been filtered to remove noise, recalibrated to ensure absolute mass accuracy, and curated to provide quality and consistency. All manually curated MS/MS spectra are fully annotated with the substructure of key ions (Figure 11 inset). Using the mzLogic algorithm, these structures can be used to identify unknowns for compounds with structures similar to, but not in, the mzCloud mass spectral database. The library's contents are updated weekly to provide broad chemical coverage and to ensure relevance to all small-molecule applications.

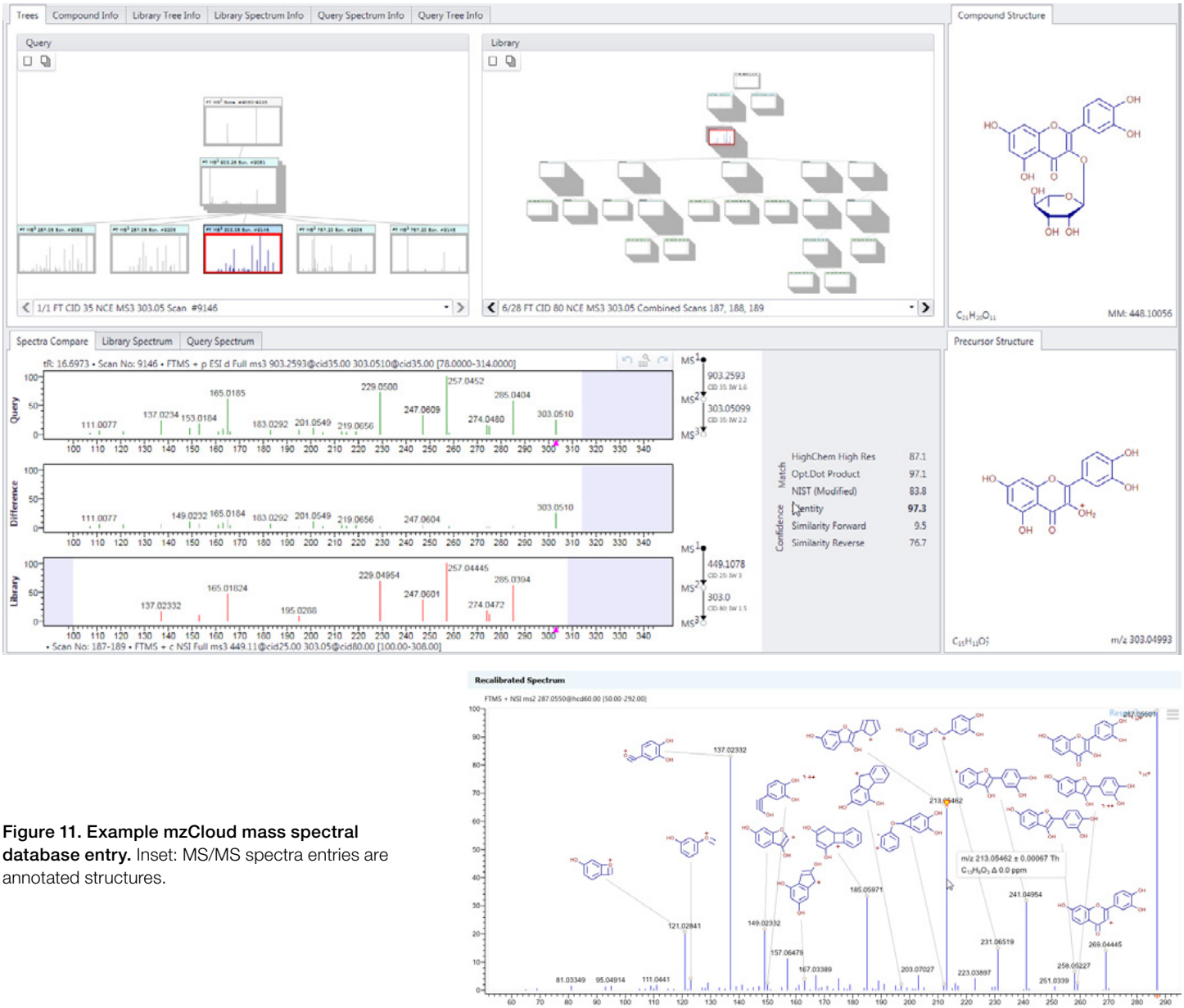

Figure 11. Example mzCloud mass spectral database entry. Inset: MS/MS spectra entries are annotated structures.

When the mass spectral database doesn't provide a match, how can the mzLogic algorithm help me identify unknowns?

Given the extent and diversity of small molecule chemical space, it's not unusual when there's no good spectral library match. With HRAM data, it's possible to determine a molecular formula to use in a database search, for example ChemSpider, the human metabolism database, KEGG, or BioCyc, but these searches often produce tens to hundreds of hits for a particular formula.

The mzLogic algorithm is uniquely able to solve this problem using similarity searching. Shown in Figure 12, the algorithm takes a HRAM spectrum with its fine-isotopic information, calculates an elemental composition, and searches a specified database of structures—for example ChemSpider—to find putative structures. The algorithm then searches the extensive fragmentation information within the mzCloud mass spectral database to find similar structures and generates a ranked list of proposed compounds based on structural similarity (Figure 13). In addition to mzLogic which is based on actual fragmentation data, the FISh scoring algorithm (powered by Mass Frontier software) can be used to perform in-silico fragmentation of the structural candidates. The result is an annotated fragmentation spectrum as well as a score that indicates how many fragments can be explained by the algorithm (Figure 14). This score can also be used to rank structural candidates in a similar fashion to mzLogic or to confirm the results obtained by mzLogic.

Even though the user can't always conclude that they have absolutely identified the component, the algorithm provides a very good clue. To be certain, the user could either synthesize or buy a standard to run to compare against the unknown data.

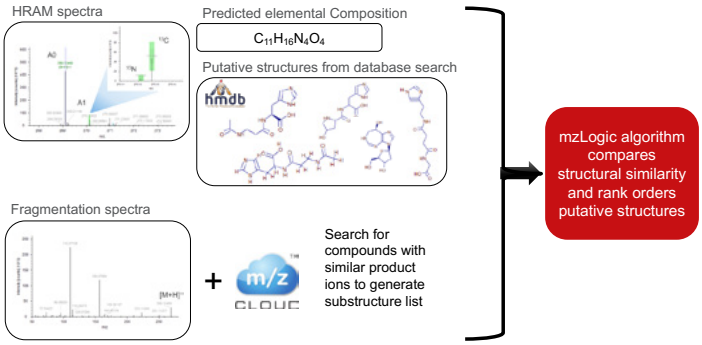

Figure 12. mzLogic algorithm similarity search approach.

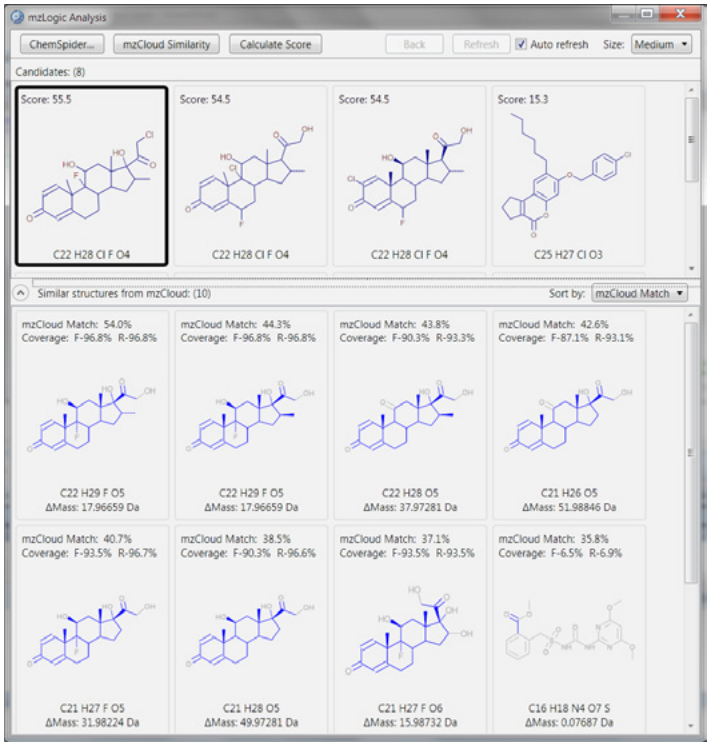

Figure 13. The mzLogic algorithm ranks similarity search results. Here, several similar structures were found that point to a corticosteroid.

How can I quickly view the chemical relationships in my sample components to enhance unknown identification?

Compound Discoverer software enhances visualization of chemical relationships in the Molecular Networks View (Figure 15.) Each of the colored dots in the view represents a sample component. The green dots indicate identified compounds and the blue dots unknowns with predicted elemental composition. The size of the dot represents the component intensity. The software uses the relationships between MS/MS data, as well as known transformations like the loss of water or acetylation, to build the constellations.

The view permits searching the data, for example by transformations or by compound name. Matches are displayed as red dots. The data shown in Figure 14 were collected in an experiment to investigate compounds that leach from O-rings such as PEG polymers. When “PEG” is used as a search term, PEG components are highlighted. When the user hovers the mouse over a dot, the software displays the compound identified (if identified) along with its information.

Users can probe constellations to view the relationships between identified and not identified components to make identifications. In Figure 15, many PEG components were identified, but in the constellation circled in red, there is an unknown component that is somehow related to identified components. When the user clicks on the connections between components in the constellation, the reaction is displayed, as well as the spectral similarity score (Figure 15).

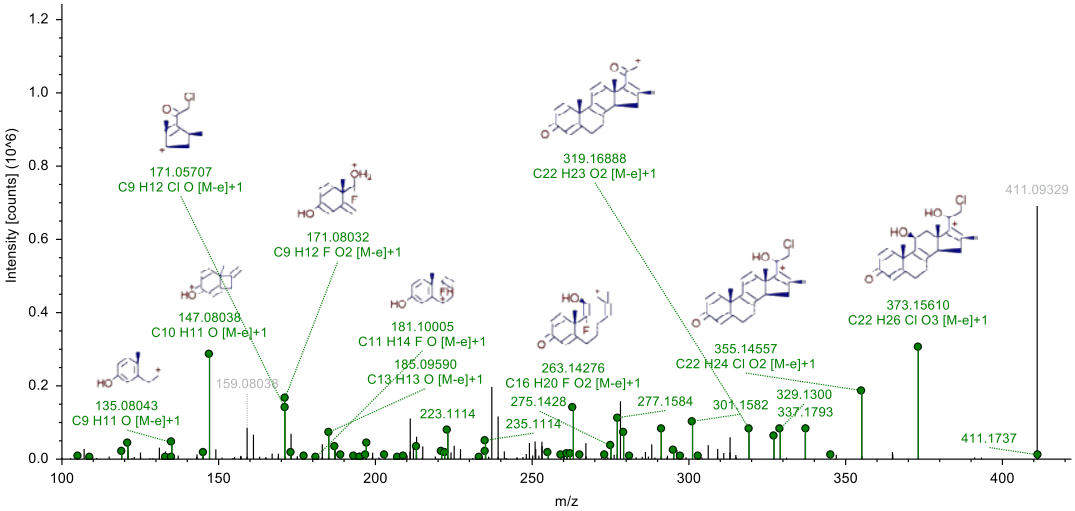

Figure 14. FISh scoring and –annotation, shown for the first structure (Clobetasol) from the mzLogic results in Figure 13, confirming the putative identification.

Figure 15. Identification of unknown components using reaction information in the Molecular Networks View. When connection 1 between the known and unknown component is clicked, the software displays the spectral match scores and the chemical reaction: the addition of C_2H_4. Continuing, connection 2 shows the addition of O_2, and connection 3, the addition of C_2H_4O.

How can I easily place identified components into their biological or process context?

Compound Discoverer software includes powerful capabilities for mapping results onto biological and process pathways, including the Kyoto Encyclopedia of Genes and Genomes (KEGG, the BioCyc collection of Pathway/Genome Databases (PGDBs) and Thermo Scientific™ Metabolika™ pathways. Figure 16 shows color-coded identified components placed on a KEGG pathway. BioCyc provides additional options to overlay pathway information from multiple samples. However, it is the Metabolika pathway tool that provides the most flexibility by permitting users to: 1) edit its default pathways, 2) build custom metabolic and biotransformation pathways and other processes, and 3) color-code samples in groups of samples as up- or down-regulated.

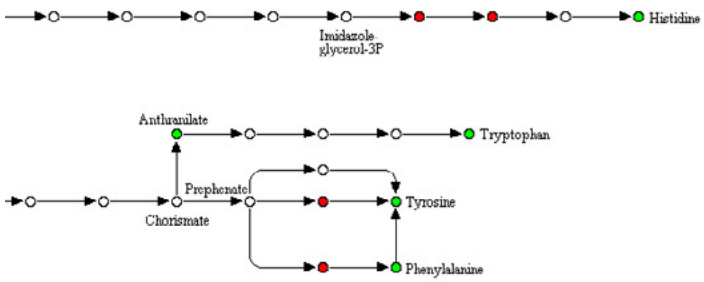

Figure 16. A KEGG pathway that is color-coded to show compounds that were identified using formula or exact mass (red) or spectral match using mzCloud (green).

What tools can I use to streamline stable isotope labeling experiments?

Compound Discoverer software includes a complete workflow to track labeled and unlabeled compounds for subsequent visualization on metabolic pathways.

All the user needs to do is provide the label and which are the labeled study samples and the unlabeled reference samples. The software goes through the data acquired to find all the components in the unlabeled samples and their elemental composition. The software then goes back through the data to find the related labeled compounds (isotopologues) based on the formulas of unlabeled components. In this manner, the software is able to perform untargeted, stable isotope label detection with very little user input. Ways to visualize results include isotopologue distribution charts and exchange rates (label incorporation rate) for each isotopologue in each sample (Figure 17). Results can also be displayed on Metabolika pathways and labeled with average or relative exchange rates.

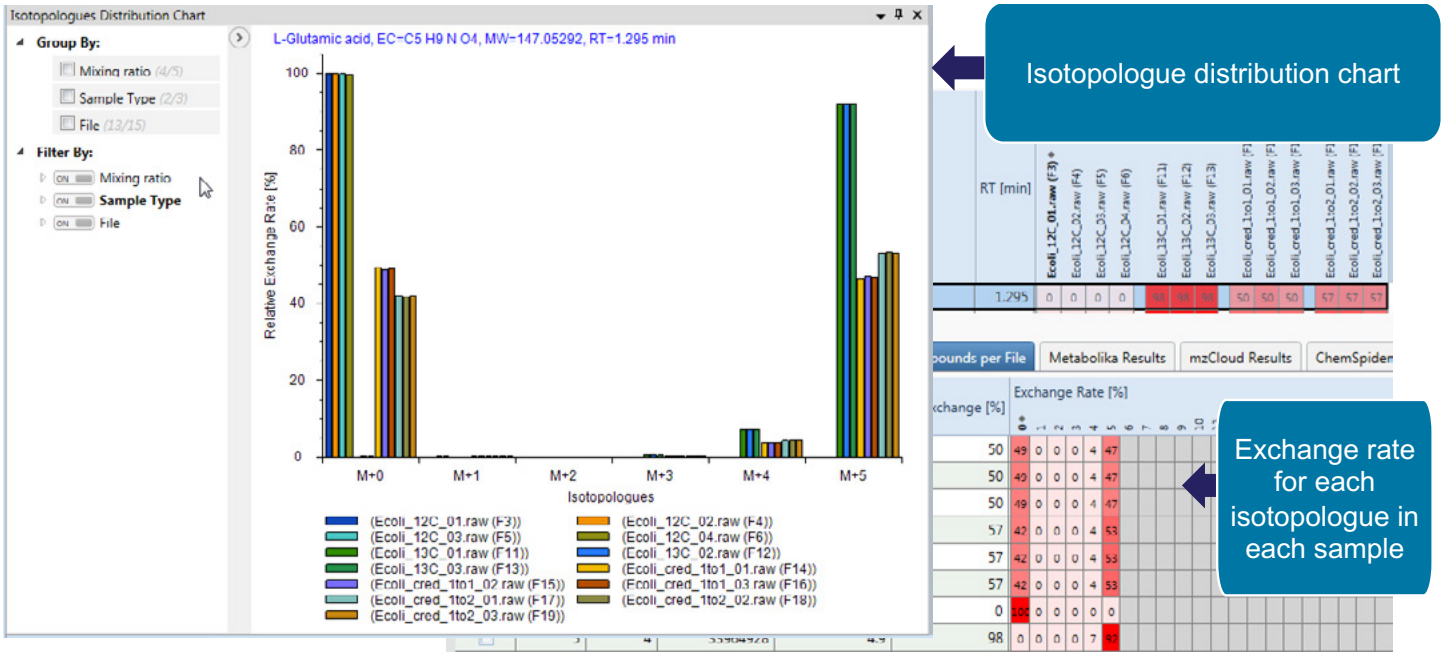

Figure 17. Isotopologue distribution chart and exchange rate plot.

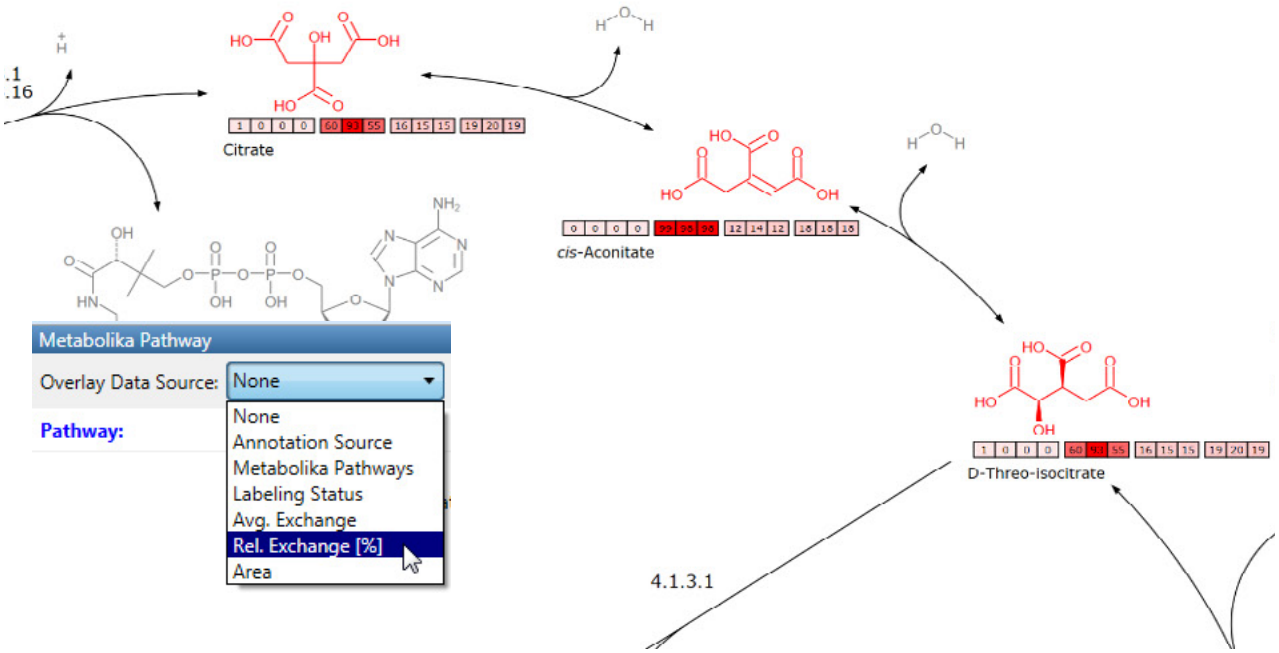

Figure 18. Stable isotope labeling experiment results displayed on Metabolika pathways and labeled with relative exchange rates. In this particular experiment four groups of samples were analyzed, unlabeled as well as three mixtures in different ratios of labeled and unlabeled sample.

How can I learn more about Compound Discover software?

mycompounddiscoverer.com provides numerous training videos, tips and tricks, and a help button to request help with a particular analysis. The website also keeps users up to date with the latest Compound Discoverer software developments.

thermofisher.com/compounddiscoverer provides more information about the cross-market applicability of the software, as well as more information about the relevant features and functionalities contained within this Smart Note.

Find out more at
thermofisher.com/compounddiscoverer

ThermoFisher SCIENTIFIC