

# Rapid Metabolite Identification Using Advanced Algorithms for Mass Spectral Interpretation

Mark A. Bayliss; Margaret Antler;  
Graham McGibbon; Vitaly Lashin

ACD/Labs, Toronto, CANADA



Advanced  
Chemistry  
Development

ACD/Labs

## NOVEL ASPECT

Application of multiple redundancy approach to extraction of relevant LC/MS features from metabolite identification studies.

## INTRODUCTION

Recognizing differences between related LC/MS data sets is the basic premise for the determination of potential metabolites in drug development. Finding small differences between two or more datasets requires a deep and rigorous analysis of each data set to extract and determine those m/z values that give rise to chromatographic peaks. The major challenge is that the signal-to-noise ratio decreases as the limit of detection is approached so the number of false positive peaks that populate the output increases exponentially. In searching for relevant and important potential metabolites, it is critical that as many false positive potential metabolite features as possible are eliminated from the output to reduce the human investment of time and energy in reviewing the results. The objective of this work is to examine and demonstrate the effectiveness and value of using software to investigate LC/MS data post-acquisition for metabolites using a chemometric approach for the extraction of the metabolites

## METHODS

### Data Acquisition

LC/MS data were previously acquired with a system consisting of an Agilent 1100 Series LC with a thermostatic well plate sampler, binary pump, and thermostatic column compartment; and an Agilent 1100 Series LC/MSD Trap XCT Plus ion trap mass spectrometer. The LC/MSD Trap XCT Plus was controlled by version 5.2 software. Chemicals were used as obtained from Sigma-Aldrich (St. Louis, MO, USA). S9 liver homogenate from a single Sprague-Dawley male rat was purchased from In Vitro Technologies (Baltimore, MD, USA) and stored at -70°C until use. Metoprolol was incubated with agitation in the presence of S9 liver homogenate for one hour. Simultaneously, solutions of the drug (30 µM) were agitated at 37°C for one hour in 1.0 mL of 2% sodium bicarbonate containing S9 liver homogenate (1.0 mg/mL) with (Metabolized) or without (Control) co-factors G6P (6.9 mM), NADP (550 µM), and G6PDH (1.5 units/mL). Reactions were quenched with 100 µL cold perchloric acid, and then centrifuged (10,000 RPM, 15 min) to pellet the proteins. The aqueous layer was injected into the LC/MS system.

### Data Analysis

Data Analysis was performed using ACD/MS Manager Suite, version 10.06, with the ACD/IntelliXtract™ add-on and an additional prototype version (V11.0) which includes a new dynamic isotope filter.

IntelliXtract is an algorithm for componentization of LC/MS data. This software, which we have developed, uses extracted ion chromatogram peaks and as many mass spectral identifiers as possible to ensure that all features relevant to a chemical species are identified. We assume that for a mass spectral isotope cluster to be relevant, it should contain at least two or more contributing isotopes which share the same chromatographic elution profile and peak characteristics. Further, when considering the isotopes within a cluster, the ratio of contributions of the <sup>12</sup>C and <sup>13</sup>C are calculated and assigned a pass or fail criteria depending on their intensity contributions relative to an estimated number of carbon atoms. During the data analysis, as mass values are characterized, the components are attributed with MS relevant "peak tags" in categories identified as <sup>12</sup>C, <sup>13</sup>C, <sup>12</sup>C+2, <sup>12</sup>C+3, <sup>12</sup>C:<sup>13</sup>C=Pass/Fail, A+2 elemental contributions. These "peak tags" are then available for the analyst to dynamically filter the results to a manageable number of appropriate and relevant component entries.

## RESULTS and DISCUSSION

We denote a single chemical entity eluting in chromatographic space as a component. Componentization is therefore defined as the process of determination of which individual ions are associated with a component, based on ion chromatographic peak profiles. Componentization requires a combination of chemometric and mass spectrometry knowledge. It is a function of sensitive chromatographic peak extraction, discriminating peaks from noise, and determining which peaks belong to the same component. Componentization is also a function of ion mass cluster determination, including isotopic ions, and analyzing adduct ions, multimer ions, and possible fragment ions. Parameters that accounted for these factors were incorporated into the software and can be selectively modified by the analyst. As a result, components are annotated with information that can be classified as peak tags.

## 1. Componentization

### Extracting the Relevant Ion Chromatograms

Figure 1a illustrates that even for relatively simple LC/MS datasets, the total ion chromatogram (TIC) may only be of use for observing ionized molecules from the parent compound and perhaps a few of its major metabolites. In this example, the largest peak is Metoprolol, a beta-blocker, and the other peaks corresponding to phase I type metabolites. With knowledge of common biotransformations, one can extract ion chromatograms of anticipated metabolites straightforwardly, however, recognizing trace components, and/or components that appear at unexpected m/z values, is more laborious. We also note that just because there is a chromatographic peak at an expected m/z value, this doesn't necessarily mean that this is the [M+H]<sup>+</sup> or [M-H]<sup>-</sup>; and mass spectrum interpretation is necessary to assign the protonated or deprotonated molecule.

The unique components illustrated in Figure 1b include the major and minor metabolites of Metoprolol at m/z 195.13 (desmethyl deamination), 226.13 (desisopropyl), 254.19 (desmethyl), 268.13 (ether to acid), 284.14 (hydroxy), and 300.13 (dihydroxy). These components were extracted automatically using IntelliXtract.

Figure 1c shows a 2D projection of the LC/MS data that provides a bird's-eye view of the sample. We zoom in for closer inspection of the expected isotopic peaks within the major metoprolol component. For less abundant ions the heavier isotopomer, primarily <sup>13</sup>C, can be challenging to confirm without examining the individual spectra in magnified detail. Isotopes are considered explicitly by the IntelliXtract process as described below.

### Peak Tags

Figure 2 shows the Table of Mass Chromatograms for components automatically extracted by IntelliXtract. During componentization, IntelliXtract determines which individual ions combine to form a component, and interprets the 'pure component' mass spectrum, assigning <sup>12</sup>C, <sup>13</sup>C, or <sup>12</sup>C+2 within each ion cluster, then identifying [M+H]<sup>+</sup> or [M-H]<sup>-</sup>; adducts, multimers, and possible fragment ions. IntelliXtract adds peak tags for each of these mass spectral identifiers, and also considers the presence and relative abundance of certain isotopomers, which are denoted as <sup>12</sup>C and <sup>12</sup>C:<sup>13</sup>C in Figure 2. A 'Pass' for the latter indicates that the presence of co-eluting isotopes was recognized and the ratio was within acceptable limits, whereas a 'Fail', outlined in red boxes in the figure, indicates that the <sup>12</sup>C/<sup>13</sup>C is outside of the expected range, and may require manual review. The software allows further filtering and sorting of the Table according to peak tag information. This facilitates the identification of real components and minimizes the number of false positive features that needed to be examined. For metabolite identification, the resultant lists of potential components was further reduced by comparing known metabolic mass differences from the parent mass.

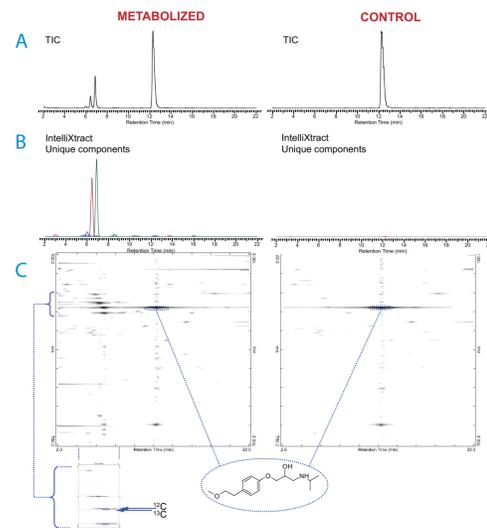


Figure 1: Comparison of LC/MS datasets from metabolized and control samples of Metoprolol that show total ion chromatograms (TIC), IntelliXtract extracted ion chromatograms for metabolites m/z 284, 254, 195 (Unique Components), and LC/MS 3D-MAPS. Blow-up region shows isotope peaks of major metabolites. Original data courtesy of PD Perkins, Agilent Technologies.

## Intelligent Recognition of Metabolic Mass Differences

The green box in Figure 2 shows how IntelliXtract labeled certain components as Potential Modifications, based on a list of known biotransformations. The software adds another peak tag corresponding to the potential modification, which may then be used to sort the table and extract the components with [M+H]<sup>+</sup> values corresponding to predicted metabolites. Other individual components with the same [M+H]<sup>+</sup> values, but different retention times, can be seen distinctly or grouped together. Components that are not originally annotated may still represent potential biotransformation products, for instance the [M+H]<sup>+</sup> 195 desmethyl deamination product.

## 2. IntelliXtract for Comparison of Samples: Metabolized vs. Control Uniqueness Tags

It is standard practice in metabolite identification to obtain data from a control sample to avoid false positive assignments. Figure 1 provides a comparison of the total ion chromatograms (TIC), identified possible metabolite ion chromatograms (Unique Components), and contoured 3D-MAP representations for the LC/MS data from the Metabolized and Control samples. Evident from all of these are new features in the Metabolized sample that could represent metabolites. Also obvious is that IntelliXtract XICs reveal even minor components that are not similar between the samples. Components are classified as similar, different, or unique compared to the Control. Figure 2 shows that known and potential new metabolite features were retained by filtering using this 'Uniqueness' tag to leave only components unique to the Metabolized sample. Criteria for Uniqueness are both chromatographic and mass spectral.

Figure 2: Table of Mass Chromatograms after IntelliXtract processing of 'Metabolized' sample dataset with Peak Tag filtering for monoisotopic parent ions.

### Compare Components Based on Chromatographic Properties (tR, Area)

Figure 2 shows that chromatographic features assigned as components by IntelliXtract have characteristic retention times and peak areas which are annotated in the Table of Mass Chromatograms. Differences in these values are recognized in the compare process and the values themselves are used to filter and sort the data. Components tagged as Unique were not present in the control sample. There are several unique components in this data set which were not identified as a potential metabolite based on expected biotransformations. Further investigation of these components, and their mass spectral characteristics is warranted.

### Compare Components Based on Mass Spectrometric Properties (m/z, Isotopes)

The isotope pattern for each ion cluster within the component was also compared by the algorithm. This can be useful for comparing components which are classified as Different or Similar based on chromatographic peak areas. The output of this test is shown in Figure 2, in the right-most column of the table. Ion clusters within a component with similar patterns between the sample and the control will pass the isotope pattern test, and those that differ will fail. With this data set, we were able to confirm that components marked as Similar did have similar mass spectral characteristics, and therefore, could not further reduce the list of potential metabolites.

## 3. Data Review—Additional Filters and Labels

### Mass Chromatogram Table Filters—Using Peak Tags

In Figure 2, each component is tagged with multiple pieces of LC- and MS-specific information that provides reviewers of the data with important information about the nature of each mass within the spectrum. The comparison with a control, <sup>12</sup>C:<sup>13</sup>C ratio and other isotope pattern information, are particularly important for reducing the number of false positive identifications without discarding relevant information (i.e., false negatives), even for ultra-trace components. The chromatograms of components with peak tags indicating <sup>13</sup>C and other features in the data aren't listed in the table, which makes initial review of the data more rapid. Analysts can explore in further detail by simple toggling of the dynamic filters to include components with a variety of custom peak tags as depicted in Figure 3.



Figure 3: IntelliXtract advanced filter option examples

### Isotope Ratio Filter

Many metabolite profiling experiments for drug development rely on the use of radiolabeled compounds to characterize the products and benefit from a comparison of different types of chromatographic data as shown in Figure 4. We have recently developed a prototype filter for non-natural isotope abundances that can be applied to find radiolabeled metabolites in datasets generated from these types of samples. One example is for a sample containing <sup>14</sup>C-Buspirone. The use of the novel isotope filter with appropriate <sup>13</sup>C:<sup>14</sup>C ratios provided a short list of priority components, including expected and unexpected metabolites with only one false positive result after the application of the filter. All metabolites known to be present in this dataset were extracted by the Isotope Ratio Filter, without the use of the control sample as a comparison. The Isotope Ratio Filter is not only useful for radiolabeled samples; it is also useful for extracting chlorinated, brominated, or sulfonated compounds based on their characteristic isotope patterns.

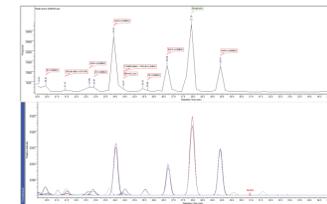


Figure 4: The top trace is the annotated radiochromatogram; the bottom trace is the aligned ion chromatograms for components that remained following extraction and filtering of the raw data, and the application of the isotope filter which was set to pattern match for <sup>13</sup>C:<sup>14</sup>C in a 1:1 ratio +/- 20% relative intensity. Note the almost 1:1 relationship of the two traces, including a number of trace level and co-eluting metabolites.

## CONCLUSIONS

Datasets containing thousands of extracted ion chromatogram peaks were reduced to a handful of relevant chromatographic components using automated software processing and peak tag filtering.

Using the multiple information redundancy approach to data extraction, we determined with good accuracy the important and relevant components within a series of datasets. This was achieved by applying peak tags based on mass spectral characteristics to the data, and a combination of filters to view the results.

The use of a wide range of spectral characteristics, such as the <sup>12</sup>C:<sup>13</sup>C ratio, and the [M+H]<sup>+</sup>, allowed us to extract a short list of components to manually review. A list of known biotransformations allowed us to quickly extract components which are potential metabolites.

The use of IntelliXtract for comparison of samples based on chromatographic peak areas and isotope patterns allowed us further reduce the list of potential metabolites to review. Both expected and unexpected metabolites were extracted with this approach.

The Isotope Ratio Filter was able to quickly and reliably extract radiolabeled components based on a user-defined <sup>13</sup>C:<sup>14</sup>C pattern, even for ultra-trace components, with only one false-positive identification, and zero missed metabolite components.

The use of spectral characteristics and their corresponding peak tags allowed us selectively filter the components for fast and easy extraction of potential metabolites from LC/MS data.

## References

- Kerns, E. H. et al.; J. Chromatography B. **1997**, 698, 133-145.
- Zhu M. et al.; Drug Metabolism and Disposition **2005**, 33, 500-507.



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