Identification of Small Molecules via Real-Time Library Search on an Orbitrap Tribrid Mass Spectrometer

William D. Barshop, Jesse D. Canterbury, Brandon J. Bills, Vlad Zabrouskov, Seema Sharma, Thermo Fisher Scientific, 355 River Oaks Parkway, San Jose, California, United States, 95134

ABSTRACT

Purpose: To enable the spectral library search of MS[2] scans concurrent with instrument acquisition to enable complex decision-making based on search scores for the identification of small molecule analytes.

Methods: We conceived of and implemented a version of the Tune instrument control software capable of performing a spectral library search in parallel to data acquisition. Version 3.5 of the Tune software allows ingestion of mzVault™ format spectral libraries (.db), like the downloadable mzCloud™ libraries, to match against experimental MS[2] scans within the "Real-Time Library Search" (RTLS) filter. For each scan, both a cosine similarity score and a confidence score are provided as outputs for thresholding and decision-making through the user dictated design of the method.

Results: Both the Thermo Scientific Orbitrap IQ-X and Eclipse Tribrid mass spectrometers are suited to leverage the RTLS infrastructure within the Small Molecule application mode, as implemented and made available in the version 3.5 release of the Tune instrument control software.

INTRODUCTION

The use of in-acquisition decision-making via real time identification of analytes by mass spectrometry has largely been focused on enhancing quantitation for peptidic compounds labeled with isobaric tags[1]. Here, we explore changes to the instrument control software of the Thermo Scientific™ Orbitrap Fusion™ Eclipse™ and IQ-X™ Tribrid™ mass spectrometers to enable the real-time identification of small molecule analytes via spectral library search. These identifications can be used as a filter or trigger to control complex decision trees in custom defined methods, for example, to interrogate specific compounds by MS[n] or alternative activation modes or energies.

MATERIALS AND METHODS

Software Development

The Real-Time Library Search infrastructure shares a common underpinning with the peptide-focused Real-Time Search (RTS) implementation. Together, the RTS and RTLS shared service is referred to here as the "Windows service" or "Real-Time Search service", which runs as a C# companion application to the instrument control software.

PRIMARY CHALLENGES

- To enable comparisons of experimental spectra to library spectra at speeds compatible with instrument scan generation.
- To leverage score outputs of RTLS to make guide decisions during acquisition.
- To extend the use of Real-Time Search has beyond the bounds of proteomics, into the small molecule realm.

SEARCH OVERVIEW

- The Real-Time Search infrastructure was modified to enable library search functionalities.
- RTLS can use both custom mzVault libraries, or local copies of the mzCloud MS[2] libraries in the mzVault .db format.
- ddMS[2] scans are sent from the instrument to the search service, and the scored results are returned to the instrument filter.

mzCloud Spectral Library Support

mzVault Search Backend

Cosine and Confidence scores

RTLS IMPLEMENTATION DETAILS

Implementation of Real-Time Library Search

Previously, the Orbitrap Eclipse instrument has demonstrated the value of real-time identification of peptide precursors for the purpose of enhanced SPS-MS[3] fragment ion selection and quantitation. These functionalities depend on the predictable fragmentation of polypeptides as they are identified by means of the Comet peptide database search algorithm[2]. Small molecules, however, do not benefit from such predictable generation of fragment ions and are therefore most often identified by means of fragment ion relative-intensity conscious approaches like spectral library searching in which the reference library contents are often composed of experimentally acquired fragmentation spectra of purified or confidently identified compounds.
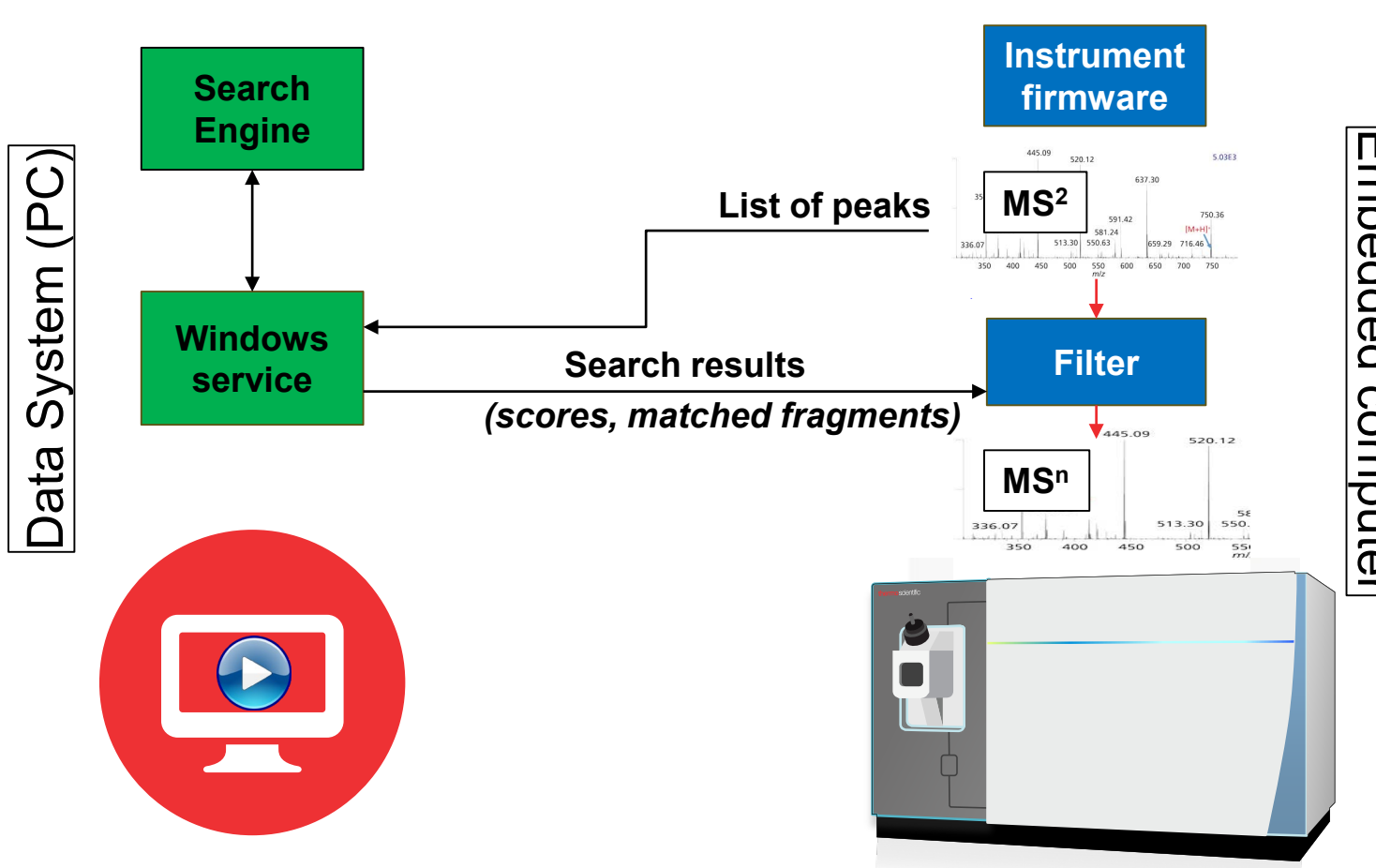

Figure 1. The Real-Time Search service receives each MS[2] scan generated during acquisition. In turn, it produces and returns search results based on user provided filter parameters, and the design of the method. Upon receipt of the search results, the instrument will apply the relevant method logic to determine if the subsequent scan action will be executed and which peaks may be selected. If the search results satisfy the thresholds within the respective instance of the RTLS filter, the scan may trigger the subsequent scan behavior defined within the method.
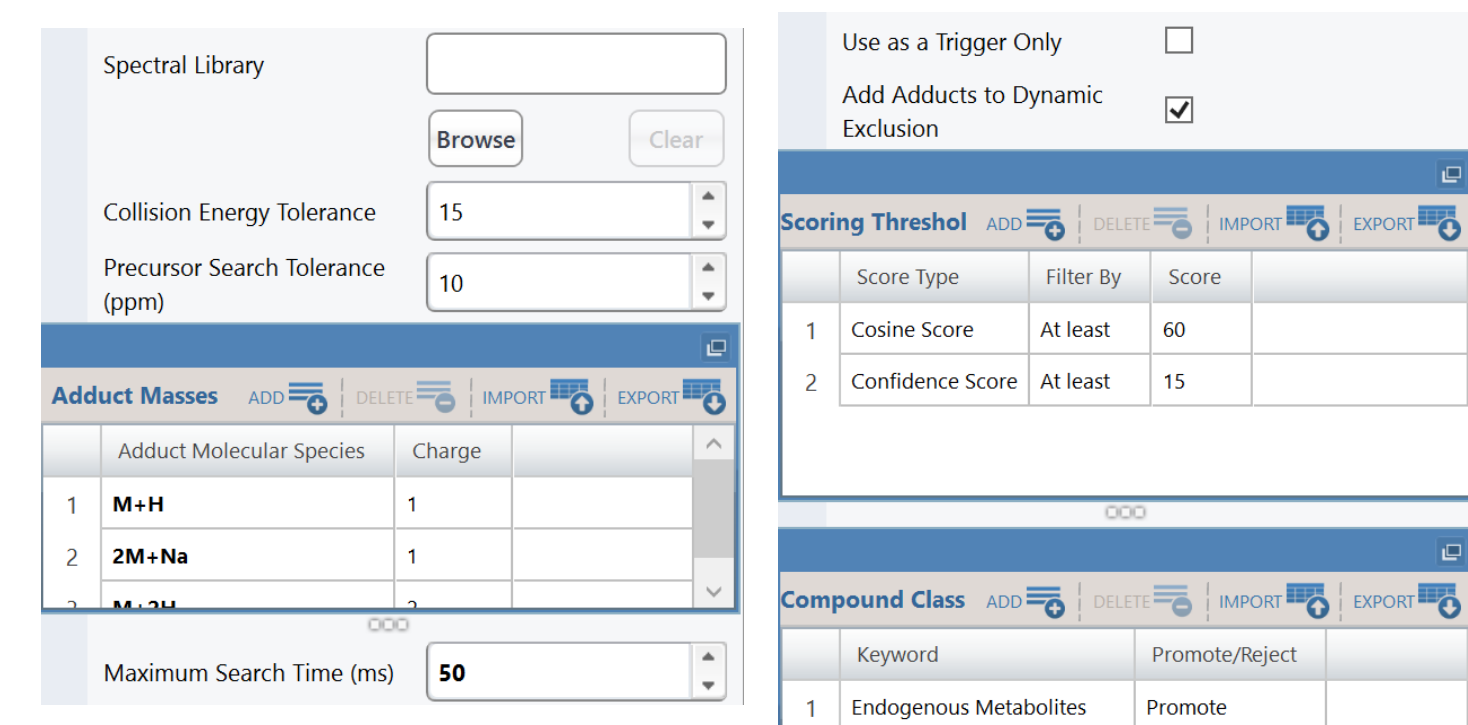
THE RTLS FILTER USER INTERFACE

Figure 2. Each instance of the RTLS filter contains two general sets of parameters: a common linked set of general parameters shared across instances (left), and a set of extended scoring and trigger parameters (right) which are unique to the instance of the filter within the method.

Search Parameters and Filter Configuration

Common parameters shared between instances of RTLS include the spectral library, collision energy tolerances, precursor search tolerance, considered adduct forms, and the maximum search duration per scan. For each filter instance, extended parameters are used to modify when and how the instrument acquires subsequent scans. Scoring thresholds may be provided to select populations either above ("At least") a score, or below ("Less than") the threshold. Fragment targets for MS[3] may be considered from any peak in the scan ("Use as a Trigger Only" enabled) or only from the population of peaks which match to the top hit library spectrum. Additionally, user specified compound classes defined within the spectral library metadata can be promoted to force passage through the RTLS filter instance regardless of scoring outputs, or rejected to ensure that the compound class is effectively ignored for downstream scan generation.
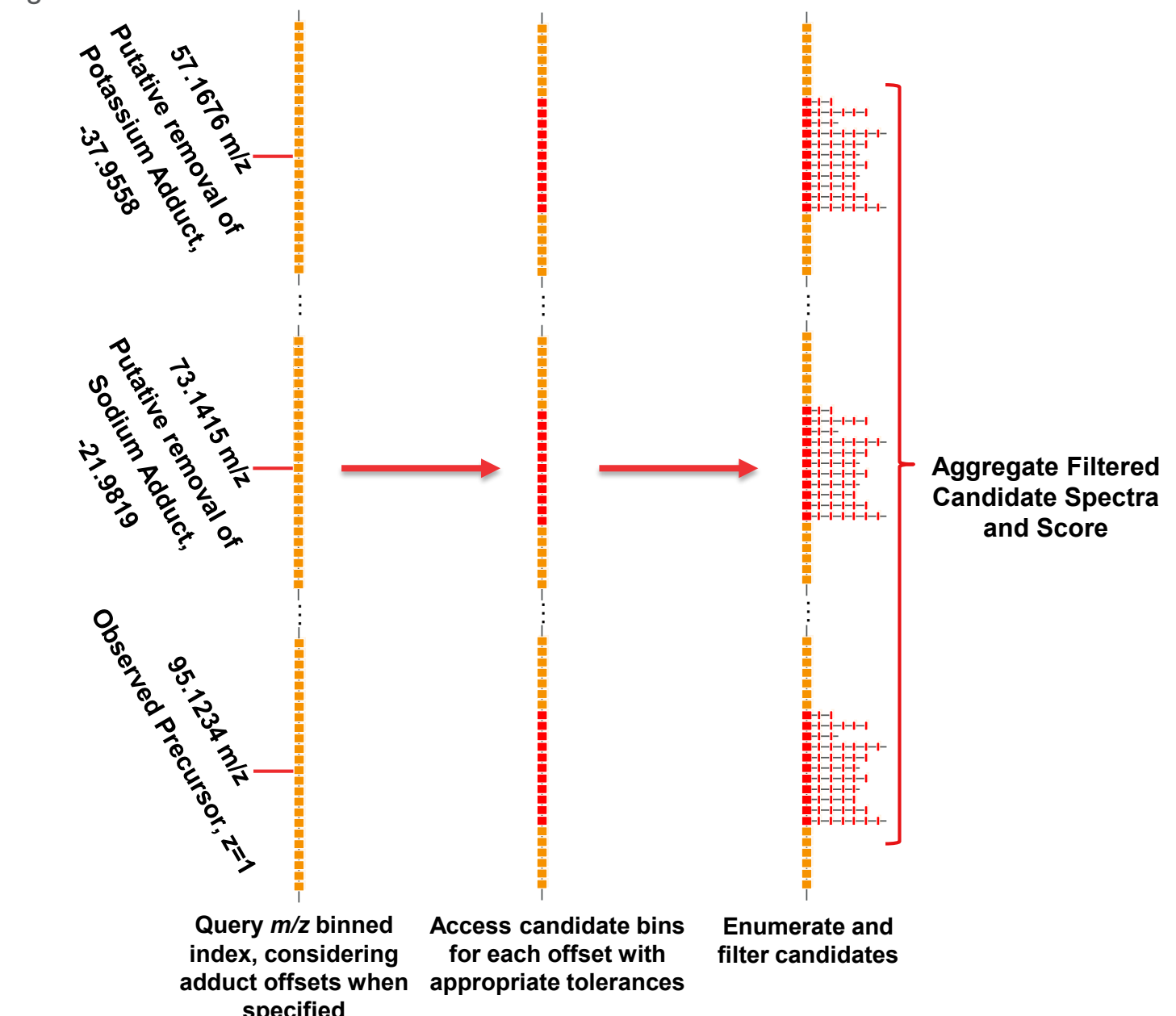

Figure 3. RTLS Library Index Adduct Offset Query and Scoring Design. If a spectral library has not been constructed to contain representative spectra of likely-to-observe adduct species, the adduct offset query functionality may be used to attempt to recover those identifications based on the expected mass offsets and can attempt to identify an analyte by its protonated or deprotonated library entry.

1. At the start of method execution, the spectral library is indexed into memory and filtered for relevant polarity, activation mode, collision energy tolerance range, and analyzer type
2. For each scan, query the spectral library at each relevant adduct offset(s) to accumulate the complete candidate spectra set
3. Candidate-Query Similarity Scoring: Generate cosine similarity scores for each selected library spectrum
4. Rank candidates by Cosine Score
5. The highest scoring spectrum from each of the top two scoring compounds are used to calculate an identification confidence score
6. Calculate Delta scores for the best match to the second best compound ID for cosine and confidence values
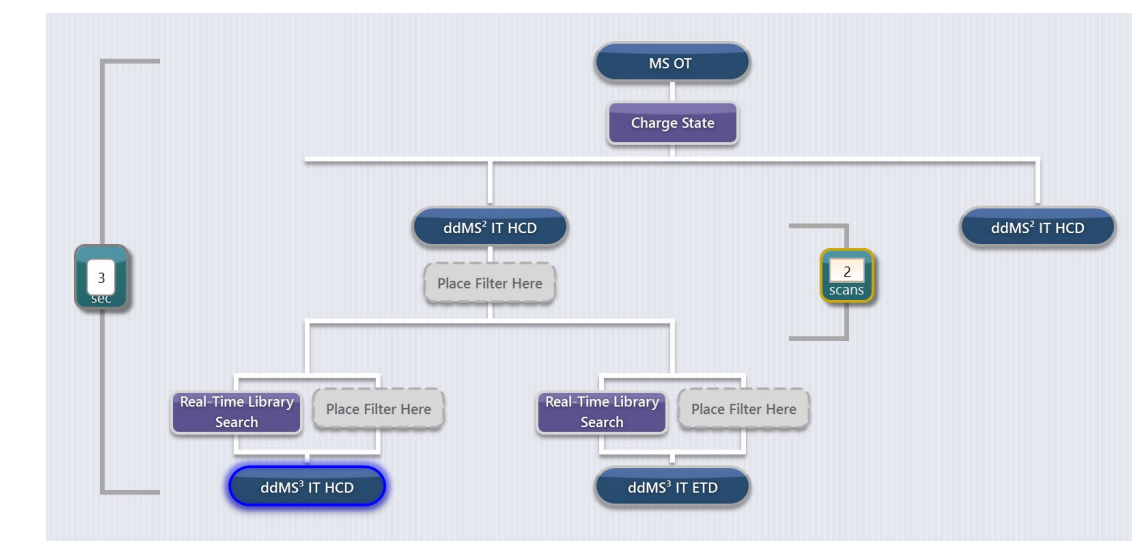
Filter Placement Constraints

Figure 4. Placement constraint of RTLS filter under a single MS[2] scan. Available locations for RTLS filters are displayed with the "Place Filter Here" notation.

Upon method execution, the Real Time Library Search Filter consumes a user supplied spectral library from disk and filters the in memory library index to retain only scans relevant based on the method architecture. Briefly, RTLS will enforce that spectral library comparisons are made against similar mass analyzers, polarity, similar activation modes, and within the user supplied collision energy tolerances based on the required library embedded metadata. While multiple copies of the RTLS filter are supported within a single method experiment, they must serve as filters to control decision-making beneath a single MS[2] scan node.

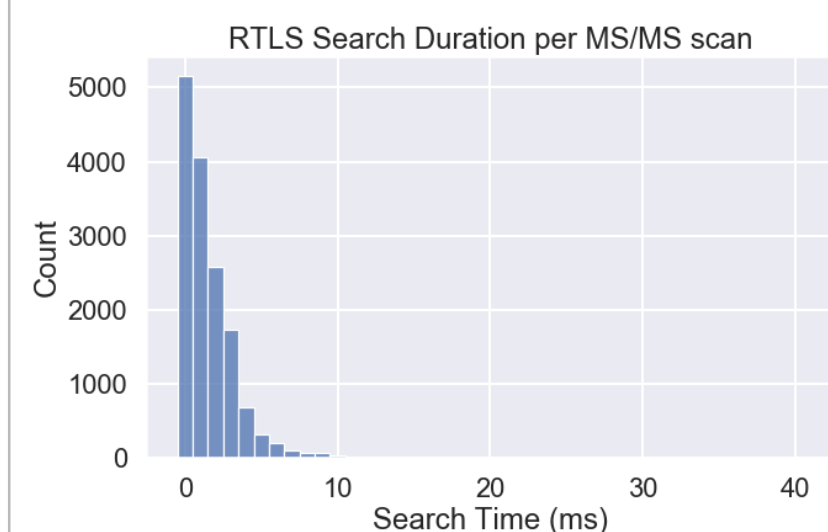### Search Duration of RTLS with mzCloud Offline Spectral Library

Figure 5. Search duration of RTLS per scan

A histogram of search durations (n = 15,130) for searches against the mzCloud Reference offline library (version 2020B), filtered for Orbitrap HCD scans between 5 and 65 NCE, and queried with a +/- 5 ppm tolerance.

SPECTRAL LIBRARY GENERATION
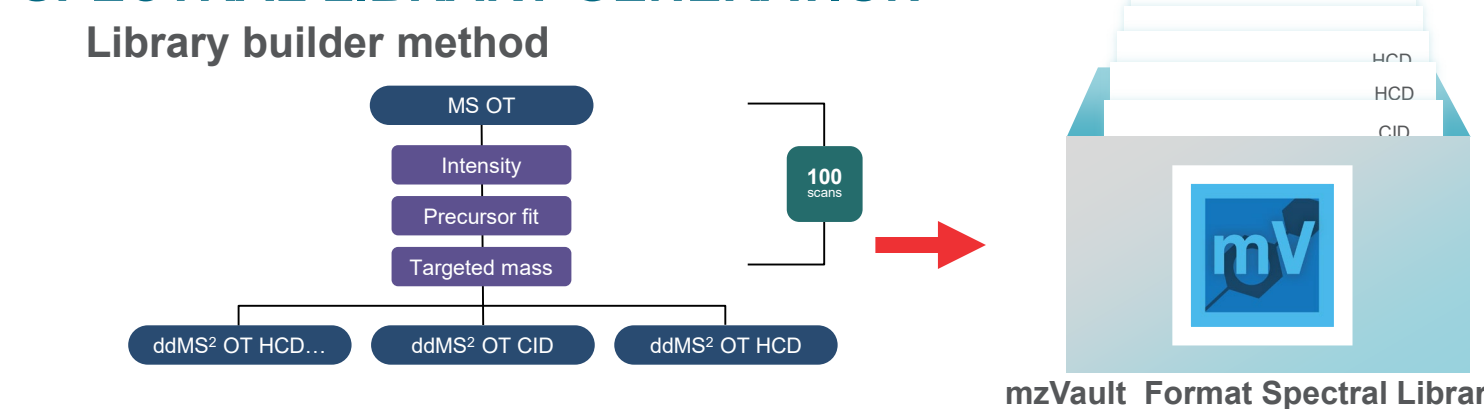
Library builder method

Figure 6. Schematic of spectral library generation via library builder template. Generally, users of Real-Time Library Search may create or obtain libraries in three ways:

1. Fragmentation spectra can be acquired for compounds of interest via the library builder method template. After acquisition, custom compound libraries are assembled within the mzVault application, directly from the raw data.
2. Existing spectral libraries may be converted to the .db mzVault format from the MassBank Record Format (*.mb).
3. Leverage downloadable copies of the mzCloud MS[2] fragmentation library, which contains over 1.7 million positive mode MS[2] spectra, and over 760,000 negative mode MS2 spectra. They are available on the Thermo Fisher Scientific Flexnet Operations Portal (Flexera, https://thermo.flexnetoperations.com/).

RTLS AUGMENTS SMALL MOLECULE ANALYSIS

RTLS
In-Acquisition Decision-making
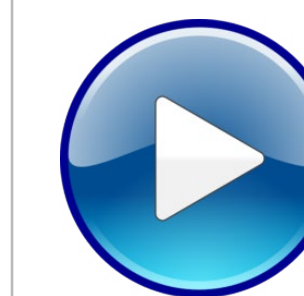
- **Library Directed MS[n]** – Target compound classes or confident/non-confident identifications with flexible threshold logic
- **Control Scan Decisions** – Multiple instances of the RTLS filter with independent settings control complex acquisition logic and downstream scans

AcquireX
Cross-Acquisition Decision-making

- **Dig Deeper** – Exclude background ions in complex matricies to enhance sampling of precursors of interest
- **Automatic Reinjection** – Comprehensive sampling of low abundance components by exclusion of previously fragmented precursors

Post-Acquisition Analysis
Comprehensive LC-MS Interrogation

- **Multi-sample Analysis** – Compile and Compare acquisitions and LC-MS feature intensities
- **Computational Workflows** – Complex treatment of acquired data, including chromatographic alignment, fragmentation tree based searching, elemental composition assignment for unknowns, pathway analyses

CONCLUSIONS

The implementation of Real Time Library Search is compatible with both custom spectral libraries, and the downloadable copies of the mzCloud spectral library in the mzVault .db format. Spectral library entries are indexed into memory during acquisition and queried for relevant comparisons with experimental spectra. Scoring outputs are generated and sent back to the instrument, at which time decisions to acquire subsequent scans can be made. Determination of acquisition behavior based on library search spectral matches facilitates powerful experiments which might include complex decision trees with multiple RTLS filters. Real-Time Library Search (RTLS) has been designed for use in small molecule applications and is available within the Tune version 3.5 instrument control software for the IQ-X and Eclipse Tribrid mass spectrometers.

REFERENCES

1. Erickson, B. K., et al. (2019). "Active Instrument Engagement Combined with a Real-Time Database Search for Improved Performance of Sample Multiplexing Workflows." J Proteome Res 18(3): 1299-1306.
2. Eng, J. K., et al. (2013). "Comet: an open-source MS/MS sequence database search tool." Proteomics 13(1): 22-24.

TRADEMARKS/LICENSING