## QqQ RAW File Parsing using Thermo’s RawFileReader SDK

As part of evaluating whether we require intermediate conversion to mzML for quantitative analysis in our SaMD, we conducted an investigation into parsing Thermo Scientific .raw files directly using ThermoFisher’s official RawFileReader SDK.

Link - <https://github.com/thermofisherlsms/RawFileReader>

This seemed to be the most official and up to date DLL that we could find. The old MSFilereader DLLs and XRawFilereader DLLs are no longer maintained.

**Aim**

The aim of this was to determine whether XCalibur raw files, particularly for QqQ data in our product can be programmatically loaded in without relying on third-party file formats or converters such as ProteoWizard’s msconvert. If these can be read into memory from the RAW file directly, we can cut out steps in conversion and therefore eliminate dependencies in our product.

### Thermo RawFileReader

We integrated the ThermoFisher.CommonCore.RawFileReader library into our C# project. The SDK provides direct access to proprietary .raw files generated on Thermo instruments. There is documentation but it is limited in what we can achieve. We managed to get it running through a Visual Studio application by adding references to the downloaded DLLs.

The first task was to extract basic functionality and we managed to get:

* instrument model
* run time
* spectrum count
* etc

This was straightforward and well-documented.

### Limitation of GetChromatogramData for QqQ SRM

The SDK includes a GetChromatogramData() function, which is suitable for full-scan MS data (e.g., full MS1 or MS2 chromatograms). However, in the case of QqQ SRM data, where transitions are acquired in a scheduled manner, this method did not return usable chromatogram information. SRM chromatograms are not precomputed in the .raw file and instead the transitions must be reconstructed from individual scans, each of which contains data for a single Q1→Q3 pair at a specific retention time.

### Structure of RAW File Data in QqQ Mode

In QqQ mode, the .raw file seems to store a sequence of MS2 scans. Each scan is acquired at a single retention time point and associated with a specific transition to the best of our understanding. This contains the data, typically with a single peak representing the intensity of the monitored Q3 ion. These scans are recorded sequentially and their metadata is accessible through scan-level filters and statistics.

### Reconstructing SRM Chromatograms

To extract complete chromatograms for each Q1→Q3 transition, Dave figured out that looping through each scan, we can iterate over all MS2 scans and extract the transition information (Q1 and Q3) from the scan filter. This retrieve’s the scan’s retention time. We can then load the data using GetSegmentedScanFromScanNumber(). For each mass-intensity pair in the data, we then store the retention time and intensity under a key for that transition. This approach allows us essentially reconstruct a time series (chromatogram) for each monitored transition directly from the raw file, without requiring conversion to mzML or relying on external tools.

Summary of our implementation:  
- identify all unique transitions in the data  
- aggregate intensity vs retention time points for each transition  
- exports each transition as a .csv file with (RT, Intensity) pairs

We should revise the export as this was only done for testing purposes and it not efficient

This approach requires only the RawFileReader DLL and no third-party dependencies and confirms that it is possible to extract QqQ chromatograms from Thermo .raw files directly using the SDK. It does however require manual reconstruction of the chromatograms.