

Supporting Information for

A Versatile Online-Offline Engine for Automated Acquisition of High-Resolution Tandem Mass Spectra

Craig D. Wenger, Michael T. Boyne II, Jonathan T. Ferguson, Dana E. Robinson, and Neil L. Kelleher*

Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

*Email: kelleher@scs.uiuc.edu, phone: (217) 244-3927, fax: (217) 244-8068

TABLE OF CONTENTS

Expanded Description of Custom Software

‘offline top-down.xls’: Offline Top-Down Identifications

‘online top-down.xls’: Online Top-Down Identifications

‘offline middle-down.xls’: Offline Middle-Down Identifications

‘online middle-down.xls’: Online Middle-Down Identifications

Software Details. Beginning with LC-MS data, the resulting file was analyzed by software developed in-house called ONLINE AUTOMATION CRAWLER. This program works with Thermo Scientific LC-MS(/MS) data files (.raw) and performed a modified version of the THRASH algorithm on every single FT scan. Chromatographic gradient information was also extracted from the .raw files and correlated with TriVersa NanoMate fraction collection times to determine the most likely plate well location for every peak detected.

After THRASH was run on all appropriate spectra, the resulting peak list was filtered on several categories, including signal-to-noise ratio, charge state, m/z range, and mass range. It is also possible to filter against inclusion and exclusion lists provided in comma-separated values (.csv) file format. There are also mass shift inclusion and exclusion lists. Mass shift inclusion lists can be used to specifically target certain modifications. Mass shift exclusion lists can be used to exclude targets that would not be desirable to specifically target for fragmentation, such as oxidations and water losses. There is also a mass shift allowed list, designed to deal with the case where a +80 Da mass shift of phosphorylation could nominally be seen as 5 oxidations. This would allow it to remain on the list of species to be targeted.

All THRASH species that remain after filtering were then grouped or “binned” by mass according to a user-defined mass tolerance in Da or ppm. This is done to ensure that species with sufficiently similar masses are grouped together, and are represented by the highest-intensity species, and ideally are only fragmented offline once.

The full data (complete output of THRASH) was then inserted into the AUTOMATION WAREHOUSE database, along with the reduced species data. Filtering and binning typically reduces the number of species detected by THRASH by 2-3 orders of magnitude. However, full data accountability is maintained, so that from the AUTOMATION WAREHOUSE database species

table, one has the ability to track from an identified or characterized protein all the way back to the original THRASH peak from a scan in the online LC-MS data file.

The AUTOMATION WAREHOUSE database functions as a de facto, intelligent high-resolution inclusion and exclusion list. High-resolution inclusion and exclusions lists are not common in modern mass spectrometers, but the binning step performed prior to the insertion of data to the AUTOMATION WAREHOUSE database makes this straightforward. The exclusion list is also far more advanced than is possible online, because the species can be marked as not necessarily just identified or unknown, but also fully or partially characterized, possibly warranting acquisition of more data in the latter case. There is no explicit exclusion list in the traditional sense, but the database works equivalently to one because the species extracted will only be those automatically recognized to be desirable targets.

In order to work with existing workflows, ONLINE AUTOMATION CRAWLER optionally allows for fragmentation scans in the online data file, although they are skipped in the analysis. These must be processed manually by the user or other software before proceeding for maximum efficiency in the automated offline stage. Identified species should be inserted into the AUTOMATION WAREHOUSE database to prevent them from being re-targeted for offline analysis. In this way, online fragmentation data can be acquired and incorporated into the data and workflow.

TARGET EXTRACTOR is a small program which was used to generate a list of species for which to acquire offline data. It connects to an existing AUTOMATION WAREHOUSE database instance and extracts a user-specified number of targets starting at a user-specified index, allowing for the user to generate a list of targets in multiple intensity intervals. This program also allows the user to specify the default fragmentation parameters for all targets extracted—by

default this is collision-induced dissociation (CID) in the ion trap with a normalized collision energy (NCE) of 35.0, activation Q of 0.50, and activation time of 50 msec. The output is an XML file that gives all information about the target and the fragmentation parameters to be applied, plus a comma-separated values (.csv) format file for easier viewing in a spreadsheet program without the fragmentation settings.

This targets XML file was then opened in MSⁿ APPLICATION, which is responsible for acquisition of offline data. The software is structured as a series of scan events for each target: isolation preview, ion trap broadband, FT broadband, isolation, and fragmentation. The user can specify virtually every setting available in the native instrument control software, and adjust these parameters in real time. Although the data acquisition is performed with the .NET wrapper library around the LTQ COM library, the output is still a normal Thermo Scientific MS/MS data file (.raw).

OFFLINE AUTOMATION CRAWLER was used to process the offline data for searching against biological databases. It uses an in-house, highly-customized version of the THRASH algorithm, since all data obtained by this platform is high-resolution FT. First, the isolation data was averaged and THRASHed to obtain the precursor mass. There is also an option in which multiple intact masses can be treated as the precursor, based on an intensity cutoff relative to the base peak of the analysis window. This allows for multiplexed fragmentation and allows more error tolerance in precursor selection, although it does also increase database search times roughly linearly with the number of precursors, assuming absolute mass searches. Fragmentation data, also FT, is then averaged and also THRASHed. The precursor and fragmentation information, as well as metadata about the particular offline event, are combined into a ProSight MS/MS

experiment, which were written into an XML-based ProSight .puf file. This file can then be searched iteratively with ProSightHT.