

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	<i>ACH B0-103</i>	Created By:	Matthew Rosin
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision #:	1.0
		Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

## Purpose

This document details how to create a compound database or a compound library using QTOF data and PCDL Manager from Agilent.

## Applicability & Authorization

BioCORE staff and BioCORE users must be trained on the Agilent 1290 LC and Agilent 6550 QTOF instrument, on Data Acquisition software, on Qualitative Analysis software, and on PCDL Manager software.

## Instrumentation & Software

Agilent 1290 UHPLC multi-sampler liquid chromatography system linked to the Agilent 6550 Quadrupole Time-Of-Flight (QTOF) mass spectrometer. MassHunter Data Acquisition (DA) software is used to control settings in both the LC and the QTOF for data collection. MassHunter Qualitative Analysis software is used to perform a preliminary investigation of the data. MassHunter PCDL Manager is used to build databases and libraries.

## Background

A Database contains only MS1 level data that is used for compound matching by looking at accurate mass and isotope ratios. A Library contains MS/MS level spectra that are used to generate a forward and reverse spectral matching score. In a library search, matching is improved by having several spectral peaks with high mass accuracy and strong signal intensity. Note that all QTOF DA methods collect MS1 level data, so a database can always be used for compound matching, while a library can only be used for MS/MS data.

Both a database and library may contain retention time (RT) information, but do not have to. Even if RT information is present in the database or library, the user is never forced to include the RT matching when performing a compound identification search. RT matching has utility for running QC samples as part of a method SOP, and for isomer identification if multiple isomers are present in a data file. For RT matching to be included, the database or library must have been created from the same LC conditions as the data file being used for compound identification.

Best practice for library creation (and the Agilent standard) is to collect MS/MS data at three collision energies (CE): 10, 20, 40 Volts. This improves library searching for compound identification. User must

**Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:**

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	<i>ACH B0-103</i>	Created By:	Matthew Rosin
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision #:	1.0
		Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

remember to extract separate MS/MS spectrum per collision energy (@ 10, 20, 40 V) in the compound finding algorithm before performing compound identification using a library.

### Reference SOPs

Users should be familiar with and may refer to the following SOPs for help in data collection and analysis:

- QTOF Tuning & Calibration\_ACHBC00008
- LC-QTOF General Maintenance\_ACHBC00010
- Qualitative Analysis of QTOF Data\_ACHBC00009
- Data Acquisition & QTOF Methods\_ACHBC00012

### PCDL Manager Software Overview

- PCDL = Personal Compound Database Library
- File extension is .cdb
- Database is populated with:
  - Compound Name
  - Mass
  - Formula
  - Retention Time (optional): enhances specificity
  - Accession Numbers (optional): CAS, Lipid Maps, Metlin, KEGG, HMDB
- Library is populated with all of the above plus:
  - MS/MS spectra
- A 1 uL injection of 100 pg/uL compound will give a very high intensity >100,000 counts of most compounds, and will give good MS/MS data

### Custom PCDL Database & Library Creation

1. Open PCDL Manager.
2. Click File> New PCDL.
3. Choose PCDL path. Best practice is to leave the default D:\MassHunter\PCDL.
4. Select Create New PCDL. Select PCDL type: General. (Also has the option to clone existing database).
5. Enter the new PCDL name and description. Click Create.
6. Click Edit button to enable editing and importing. User cannot search while editing.

**Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:**

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	<i>ACH B0-103</i>	Created By:	Matthew Rosin
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision #:	1.0
		Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

7. Click File> Import Compounds...> Select a .csv file of interest.
8. The Import may error, in which case an error message and log will pop up. These error files are saved in the location of the imported file.
  - a. Errors often result from compound name syntax. Open the error file to see what was not imported. The reason for the error will be unknown. Use trial and error to fix and re-import.
9. After import, the screen will stay blank. Turn off the Edit button.
10. Click Find Compounds button to show the imported compounds.
11. All changes in PCDL database will automatically be saved. There is no save button.
12. User can add molecular structures to compounds. Enable the Edit button and select a row to be edited.
13. Click on the Structure icon, which will open a new window. Browse to find the .MOL file. Click save.
  - a. Users can find .mol files from various sources, including: [www.ebi.ac.uk/chebi](http://www.ebi.ac.uk/chebi)
14. Subset databases can be created by highlighting compounds of interest and right clicking to Create Subset PCDL...
15. Highlighted compounds can also be added to existing databases by right clicking to Append to PCDL...
16. User can click on Advanced Search to find compounds in a database by various options.
17. Adding Spectra using .CEF file (error prone):
  - a. User can add spectra to compounds in a database using a .CEF file. This file type would be exported from Qualitative Analysis after performing Compound Finding, Identification, and quality checking.
  - b. With Edit button engaged, and a compound selected in the table, click the Spectra button.
  - c. In the pop-up box in the top right corner, click Open File and select the .CEF file containing the compound MS/MS spectra. This will populate the Acquire spectra table with details on Precursor Ion, CE, Polarity, etc. All rows will show the name of the compound that was highlighted.
  - d. User will choose the correct row including precursor ion and CE for the highlighted compound. It is easy to make a mistake here, so users should be very careful and match the Compound m/z to precursor ion (M+H).
  - e. Repeat this process for each individual compound in the library to populate MS/MS spectra.
18. Adding Spectra from Qual. (safer, but only works for named compounds):
  - a. User can add spectra to compound in a database directly from Qual. User must first run Compound Finding and Compound Identification with extraction of MS/MS data by separate CE in Qual.
  - b. If a compound is not identified by an existing database or library, but the user is confident this compound is 'real', it can be given a name like 'Unknown 1'. This name must match exactly to the compound in the Database which is the destination for sending the MS/MS spectra. User

**Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:**

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	ACH B0-103	Created By:	Matthew Rosin
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision #:	1.0
		Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

can manually change a compound name by right clicking Name in the compound list> click Add/Edit Manual Identification.

- c. Select the MS/MS product ion spectra for each named compound (matching a database compound) from the Data Navigator.
- d. Right click in the MS Spectrum window> click Send spectra to PCDL> in the pop-up window, select the Database path, min. base peak height (default is 1000 counts), All peaks, Observed m/z, and Send and replace the spectrum in the library. Note, user can choose 'Skip sending this spectrum' if the spectrum already exists in the library. Click Send.
- e. This will automatically send the MS/MS spectra to the selected database and match it with the compound with the exact same name.

#### 19. Updating or Adding RT Information to a Database or Library

- a. User can manually update RT information in the .csv file used to create the database, OR
- b. User can export a .csv file from Qual. with the compounds and updated RT information as required for updating RT, OR
- c. User can navigate to DATA (D:)> MassHunter> PCDL> Import Files> RT\_Import .csv template, and add compound information including RT as required for updating.
- d. In PCDL Manager> Import tab> Choose a .csv file to import. The left table will populate with matched compounds and the RT (Best) chosen from the import file. If 'Hits' shows more than one RT value for the compound, select the compound and the right side table will populate with the RT hits. The user can then manually choose the correct RT for that compound.
- e. When finished, click Import RT. RT information is saved automatically.

### Creating a Database and Library from Pure Compounds (Standards)

When: A user developing a targeted MS/MS method will often develop a database and library using the pure standards obtained from commercial sources. This database and library is useful for quality assurance and quality control. Research sample data files can be subjected to compound identification using the database or library to improve confidence in the presence of compounds before quantitation is performed. Databases and libraries containing RT information are very useful for running and verifying regular quality control (QC) samples as part of the method SOP. Additionally, databases and libraries with RT information are critical for confirming the identity of isomers with different RT.

1. Prepare the sample for injection for LC-MS.

Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	<i>ACH B0-103</i>	Created By:	Matthew Rosin
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision #:	1.0
		Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

2. Create a method in Data Acquisition (DA) to acquire MS1 level data in MS mode.
3. Choose some basic LC conditions (perhaps a 5-95% 'B' gradient over >15 minutes) for eluting your standards (first individually and then as a cocktail) and incorporate into the MS method.
  - a. Note: In targeted MS/MS methods the optimal LC parameters for running real biological samples will not be discovered until much later in method development. The RT information in databases and libraries can be easily updated later (see Custom Database Creation above).
4. Perform Find by Integration to extract compounds and generate a compound list.
5. User can manually input the correct name of each Standard compound by matching the m/z.
  - a. User can do this 1 compound at a time at first to ensure there is RT separation, and then run the Standard cocktail to generate a larger compound list for .csv export. Alternatively, a user can export each compound .csv and import into PCDL Manager to build the database 1 compound at a time (slower).
6. Export the compound information in the Compound List as a .csv file. This will contain the column headers recognized by PCDL Manager.
7. Import the .csv file into a new database in PCDL Manager. Although the database is now ready to be used as is, further refinement and upgrading to a Library is required for targeted MS/MS.
8. The same standard/cocktail that was used to create the database will be run again in Targeted MS/MS mode.
9. Create a method in DA to acquire MS/MS data by targeted MS/MS mode (using same LC conditions).
10. Use the compound list in .csv to enter precursor ions in the method.
11. Set the method to collect all data in three collision energies (CE): 10, 20, 40 Volts.
12. Perform Find by Targeted MS/MS and remember to extract MS/MS spectra at each CE.
13. User can export .CEF file and import into PCDL for assignment to compounds in the database.
14. The file is now considered a Library as it contains MS/MS data.

### Creating a Database and Library from Unknown Compounds in an Important Sample

When: A user may have knowledge that an important or representative biological sample will contain many compounds of interest that are desirable to deposit into a new custom database and library for further QTOF MS research. An example might be a blood plasma sample from a patient with classic symptoms of a disease with known biomarkers that is being studied by researchers, for which a database would be useful for studying further patients with classical or variants of this disease by QTOF mass spectrometry.

1. Prepare the sample for injection for LC-MS.

**Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:**

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_

Faculty/ Department or Institute/Centre	Cumming School of Medicine, Department of Pediatrics, Neurocritical Care Program, ACH BioCORE	Date Created:	January 4 2019
Location	ACH B0-103	Created By:	Matthew Rosin
		Revision #:	1.0
Supervisor	Dr. Michael Esser & Dr. David Sinasac	Revision Date:	January 4 2019
		Revised By:	Matthew Rosin

2. Create a method in Data Acquisition to acquire MS1 level data in MS mode.
3. Determine optimal LC conditions for the sample and incorporate into MS method.
4. Perform Find by Molecular Feature to extract compounds.
5. Note that because we are creating a new database from unknown compounds, there will likely be no Agilent QTOF database that contains most of these compounds. However, it doesn't hurt to look before proceeding.
6. Perform Database Search against the three (3) Metlin databases to see if any compounds are identified.
7. It is possible that some remaining compounds could be identified by online databases. Keep in mind that online databases are composed of spectra obtained from various instruments and not an Agilent QTOF, so the matching score threshold must be lower than what we would set for an Agilent library.
8. User may have to perform online database searches one compound at a time. Only some online databases allow batch searches.
9. Initially, many of the compounds in the database will not have identities. This is OK as the user will build and edit the database over time. User can use default names like Unknown 1 for now.
10. Molecular formulas should be generated for each compound using the Generate Formulas algorithm, as this will allow some annotation even if the exact identity is not known.
11. Export the compound information in the Compound List as a .csv file. This will contain the column headers recognized by PCDL Manager.
12. Import the .csv file into a new database in PCDL Manager. Although the database is now ready to be used as is, further refinement and upgrading to a Library is desirable.
13. The same sample used to create the database will be run again in Targeted MS/MS mode.
14. Create a method in DA to acquire MS/MS data by targeted MS/MS mode (using same LC conditions).
15. Use the compound list in .csv to enter precursor ions in the method.
16. Set the method to collect all data in three collision energies (CE): 10, 20, 40 Volts.
17. Perform Find by Targeted MS/MS and remember to extract MS/MS spectra at each CE.
18. User can export .CEF file and import into PCDL for assignment to compounds in the database.
19. The file is now considered a Library as it contains MS/MS data.

Printed copies are UNCONTROLLED unless signed by an authorized BioCORE lab personnel below:

Personnel Name: \_\_\_\_\_

Date Printed: \_\_\_\_\_