



MultiAlign Software

This documentation describes MultiAlign and its features. This serves as a quick guide for starting to use MultiAlign. MultiAlign comes in two forms: as a graphical user interface (GUI) or a command line console application. Both are compatible with the Windows Operating System Windows XP and 7 and for 32-bit and 64-bit platforms.

Both applications run the same backend code. Both applications produce the same output files (SQLite database and cross tab comma separated value files). The GUI application has an easy to use wizard that walks you through an analysis setup. It also has visualization capability for looking through the data and investigating results. The console application is intended for more advanced users and automated analysis pipelines. It produces a database file that the GUI can read for data visualization.

MultiAlign Software Website

Software and tutorials are provided here as of 12-11-2012:

<http://omics.pnl.gov/software/MultiAlign.php>

Test Data

Test data is provided on the MultiAlign software website. Local cache Accurate Mass and Time (AMT) tag databases are provided for peptide identification. Also provided are data for using MultiAlign through dataset only based analysis.

Windows GUI

On the MultiAlign software page (<http://omics.pnl.gov/software/MultiAlign.php>) you can find several tutorials on how to run the GUI through an analysis, or through visualization.

Console Application

The console application is described here in this quick start guide. The next sections describe how to run MultiAlign Console, its required input, and the expected output files. The file formats are described elsewhere on the MultiAlign software website.

Running the Application

The basics of the console application is that it requires the following:

1. Parameter File – describes all parameters to use for the analysis
 - a. -params
2. Input File – describes the data to be used in the analysis
 - a. -files
3. Output folder path – path where the output results should go.
 - a. -path
4. Name – name to give the analysis result database, log files, etc.
 - a. -name

Other command line options exist, that allow you to customize output, log file paths, and are accessible via the command line after typing one of the following:

MultiAlignConsole.exe -h

MultiAlignConsole.exe -help

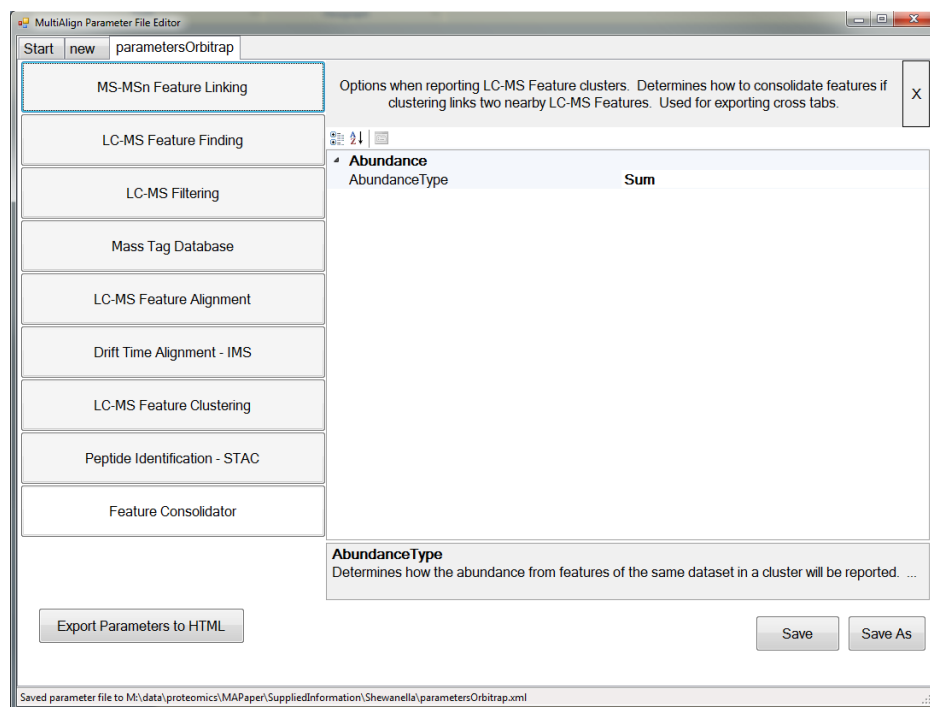
Example Command Line:

Highlighted in bold are the required command line options.

C:\Dir\MultiAlignConsole.exe **-params** parameters_16ppm_0.014NET_3ms.xml **-files** input.txt **-path** m:\data\exampleOutput **-name** myAnalysisName

Parameter Files

The parameter files are XML based files that hold each parameter for each part of the analysis process. These files are editable via text editor for the advanced user, or through MultiAlign's parameter file editor, that groups and categorizes the parameters for you. Below is a screenshot of the parameter file editor. A tutorial is provided on the MultiAlign software website.



Input Files

Input files have a specific formatting. This formatting must be correct before an analysis can be run. An input file must include the following:

1. Features Files (DeconTools, LCMSFeatureFinder)
 - a. [Files]
2. Raw Files – (Optional)
 - a. [Raw]
3. Database Input – (Optional if using AMT based approach)
 - a. [Database]

Full examples are provided on the MultiAlign software website.

Feature Files

Indicate the paths of the files you want to analyze. If you have spaces in the paths then you should use parenthesis "c:\example path with spaces\data_LCMSFeatures.txt"

Format

[Files]

M:\DMS_WorkDir7\Sarc_MS_73_21Aug10_Cheetah_10-08-04_0000_LCMSFeatures.txt

M:\DMS_WorkDir7\Sarc_MS_90_21Aug10_Cheetah_10-08-02_0000_LCMSFeatures.txt*

Required Header

The required *[Files]* is a required header. The asterisk here indicates it should be used for the alignment baseline. (See selecting an alignment baseline section for more details.) The header is not case sensitive.

RAW Data files

Indicate the paths to the raw data files (from the instrument) that NEED to associate with one of the feature files. This file names should match (minus the extensions: _isos.csv, _LCMSFeatures.txt). These are not required for input. If they are supplied then MultiAlign will extract raw data from them (MS/MS spectra) etc. for use in its algorithms.

Format

```
[Raw]
M:\DMS_WorkDir7\Sarc_MS_73_21Aug10_Cheetah_10-08-04_0000_LCMSFeatures.txt
M:\DMS_WorkDir7\Sarc_MS_90_21Aug10_Cheetah_10-08-02_0000_LCMSFeatures.txt*
```

Required Header

The required *[raw]* is a required header before the raw files are specified. Do not use an asterisk here to indicate an alignment baseline, ONLY DO THIS IN THE FEATURES FILE DEFINITION. The header is not case sensitive.

AMT tag Databases

If you want to match to an AMT tag database then you can do so with the below options (please note that some of these are only available from within the PNNL firewall. Also note that AMT tag databases can be used as the alignment baseline. Specifying a database automatically implies that you want to perform AMT tag matching.

1. **Automated Processing Engine (APE - <http://omics.pnl.gov/software/APE.php>) Cache databases** – These are local database files in the SQLite format.

```
[Database]
ape="m:\myDatabaseFile.ape"
```

2. **SQLite AMT tag databases** – These are local database files in the SQLite format

```
[Database]
sqlite="m:\myDatabase.db3"
```

3. **SQL Database from the Mass Tag System (MTS) – PNNL Only**

```
[Database]
server=myserverName
database=myDatabaseName
```

Selecting an alignment baseline

1. If you are using a dataset you can specify the baseline with an asterisk (*) as follows:

```
[Files]
M:\DMS_WorkDir7\Sarc_MS_73_21Aug10_Cheetah_10-08-04_0000_LCMSFeatures.txt
```

*M:\DMS_WorkDir7\Sarc_MS_90_21Aug10_Cheetah_10-08-02_0000_LCMSFeatures.txt**

2. If you are using AMT and have a database, you can specify as follows:

[Files]

M:\DMS_WorkDir7\Sarc_MS_73_21Aug10_Cheetah_10-08-04_0000_LCMSFeatures.txt

M:\DMS_WorkDir7\Sarc_MS_90_21Aug10_Cheetah_10-08-02_0000_LCMSFeatures.txt

[Database]

sqlite="m:\myDatabase.db3"

NOTE: You should only specify a baseline dataset in the features definition portion of the file. NEVER do so with the RAW data section using an asterisk. NEVER put an asterisks on the AMT tag database specification either.

Output Files

The output file formats are described on the MultiAlign software website. However, this section will quickly detail what you can expect:

1. Output Database
 1. Format: SQLite (<http://www.sqlite.org/fileformat.html>)
2. Abundance Cross Tab
 1. CSV formatted
 2. Each row is a cluster
 3. Each column is a value; the first columns are related to the cluster mass and NET. The subsequent clusters are abundance values only.
3. Cross Tab
 1. CSV formatted
 2. Each row is a cluster
 3. Each column is a value; the first columns belong to the cluster followed by the mass tag (if using AMT) and then each feature that was part of that cluster.
4. Plots
 1. A directory containing all of the images
5. Log
 1. A log file that