

This HOWTO contains information about the processing of multiple raw data files acquired by liquid chromatography-mass spectrometry (LC-MS) for a single Mascot (MatrixScience, London) search, followed by analysis of these files by MSQuant (CEBI, <http://msquant.sourceforge.net>). This document does not describe the way Mascot Daemon (MatrixScience) handles multiple samples for searching. MSQuant does not currently support the MatrixScience method for processing multiple samples.

---

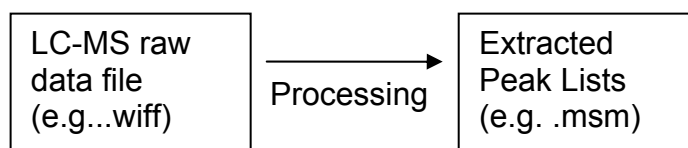
### Important points to note:

1. MSQuant is currently only able to look in individual raw data files for peptide information. It does not support the multiple samples within a single .wiff file function which is an option in Analyst QS software for QSTAR instruments. This requires that multiple sample runs be acquired as separate .wiff files.
2. This HOWTO addresses the way CEBI users have combined multiple raw data files for single Mascot searches, and subsequent analysis using our in-house developed software for protein and peptide validation – MSQuant. This HOWTO is only provided for MSQuant users interested in making use of this functionality.

## 1. Introduction

There are several reasons for wanting to fractionate samples further prior to mass spectrometric (MS) analyses. One example is to separate proteins by 1D-SDS-PAGE, dividing the entire lane into ten separate pieces and processing each gel slice for MS. This GeLC-MS (gel enhanced LC-MS) approach is frequently used in our laboratory and is the basis for this HOWTO. The use of SDS-PAGE allows a simple method to fractionate the protein sample which reduces overall sample complexity and can lead to higher sequence coverage for identified proteins.

Each sample is analyzed by a single LC-MS run, generating a raw data file which can be processed with scripts to extract MS/MS peak lists. For Mascot and ABI-Sciex QSTAR data, this is done with a script (mascot.dll).



In the GeLC-MS example, we would therefore get ten separate LC-MS raw data files (.wiff for ABI-SCIEX QSTAR, .RAW for Thermo LTQ-FT) which will be processed to generate ten peak list files (.msm files). These ten files can be concatenated to generate a single large peak list file. However, it is important to do so while properly assigning the raw file from which a peptide originates from. This is accomplished by

inserting information about the raw file within the TITLE line of a standard Mascot format peak list (which is shown below).

This is an example of a standard Mascot peak list.

```
BEGIN IONS
PEPMASS=708.450253705455
CHARGE=2+
TITLE=Elution from: 3.81 to 3.81  period: 0  experiment: 1 cycles: 1
338.905 3
356.936 4
356.974 4
374.915 4
------(truncated for brevity)-----
```

This is an example of the same peptide MS/MS peak list with appended raw data file information (in bold lettering)

```
BEGIN IONS
PEPMASS=708.450253705455
CHARGE=2+
TITLE=Elution from: 3.81 to 3.81  period: q2lcms060104_07se.wiff  experiment: 1
cycles: 1
338.905 3
356.936 4
356.974 4
374.915 4
------(truncated for brevity)-----
```

In this way, the ten raw files can be searched in a single Mascot search and the originating raw file information is visible in the pop-up tool-tip information available when you 'mouse-over' the query number (shown in figure below).

1. [IPI00022434](#) Mass: 71317 Score: 179 Queries matched: 47  
 SWISS-PROT: P02768|REFSEQ\_NP: NP\_000468|ENSEMBL: ENSP00000295897 Tax\_Id=9606 Serum albumin precursor  
 Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
<input checked="" type="checkbox"/> 262	464.167280	926.320009	926.486115	-0.166106	0	(31)	0.062	1	YLIEIAR
<input checked="" type="checkbox"/> 263	464.181360	926.348167	926.486115	-0.137948	0	(46)	0.0021	1	YLIEIAR
<input checked="" type="checkbox"/> 264	464.183107	926.351662	926.486115	-0.134453	0	(39)	0.012	1	YLIEIAR
<input checked="" type="checkbox"/> 266	464.185321	926.356089	926.486115	-0.130026	0	(46)	0.0022	1	YLIEIAR
<input checked="" type="checkbox"/> 267	464.187736	926.360920	926.486115	-0.125195	0	(38)	0.015	1	YLIEIAR
<input checked="" type="checkbox"/> 270	464.208584	926.402617	926.486115	-0.083498	0	(46)	0.0028	1	YLIEIAR
<input checked="" type="checkbox"/> 271	464.209842	926.405133	926.486115	-0.080982	0	(39)	0.015	1	YLIEIAR
<input checked="" type="checkbox"/> 272	464.213892	926.413233	926.486115	-0.072882	0	(42)	0.0073	1	YLIEIAR
<input checked="" type="checkbox"/> 273	464.215388	926.415823	926.486115	-0.070387	0	(47)	0.0055	1	YLIEIAR

Top scoring peptide matches to query 272  
 Elution from: 22.29 to 22.33 period: q2lcms060104\_07se.wiff experiment: 1 cycles: 2  
 Score greater than 33 indicates identity  
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
42.0	-0.07	1	IPI00022434	YLIEIAR
27.3	-0.11			YLIVSLR
12.5	-0.11			LVINEPSR

Top scoring peptide matches to query 271  
 Elution from: 22.75 to 22.75 period: q2lcms060104\_03se.wiff experiment: 1 cycles: 1  
 Score greater than 33 indicates identity  
 Status bar shows all hits for this peptide

Score	Delta	Hit	Protein	Peptide
39.0	-0.08	1	IPI00022434	YLIEIAR
26.1	-0.12			YLIVSLR
13.3	-0.02			YLCRCR
13.2	-0.05			SGLTEHQR
10.5	-0.09			SSTLVQHR
10.0	-0.11			LYYIDLK
8.8	-0.11			LWINEPSR
7.5	-0.16			ARVLELAR
7.5	-0.12			YLVNYKK
7.3	-0.05			YHEVTLR

<input checked="" type="checkbox"/> 652	569.764096	1137.513641	1137.490646	0.022995	0	(38)	0.019	1	CTESLVNR
<input checked="" type="checkbox"/> 653	569.770234	1137.525916	1137.490646	0.035270	0	(45)	0.0038	1	CTESLVNR
<input checked="" type="checkbox"/> 654	569.774214	1137.533876	1137.490646	0.043230	0	(45)	0.004	1	CTESLVNR

## 2. Software tools required

1. Peak list extractor capable of outputting Mascot generic flatfile peak lists (e.g. mascot.dll supplied as a processing script with Analyst QS)
2. MultiRawPrepare.pl with MultiRawPrepare.bat driver file
3. Mascot search engine (for performing the search)
4. MSQuant
5. A PERL distribution to enable the running of the PERL script MultiRawPrepare.pl (e.g. ActiveState Perl <http://www.activestate.com>)

Tools 2 and 4 are developed by CEBI.

MultiRawPrepare is available from <http://www.cebi.sdu.dk/software>

MSQuant is available from <http://msquant.sourceforge.net>

## 3. Step-by-step instructions

1. Names of extracted peak list files should contain the raw data file name within it. (for example, Q2LCMS060104\_12se\_BSA\_500\_1.msm (peak list) and Q2LCMS060104\_12se.wiff (raw data file))
2. Extracted peak list files (one for each raw data file) and raw MS data files (e.g. .wiff) should be placed in the same directory.
3. The batch driver file for MultiRawPrepare PERL script should be edited to reflect the correct directory with the 'SET INDIR=' (from point 2.)
4. The output file should be specified as well with 'SET OUTFILE='.
5. The MultiRawPrepare.bat should be run, the script will detect the raw files and peak list files, then concatenate the raw peak list files into the single specified output file (from point 4.)
6. Mascot search should be initiated with the combined output peak list file.
7. Mascot result file (Peptide summary) should be saved as HTML in the same directory as raw data files.
8. MSQuant should be directed to look within this directory for the combined HTML file and the raw LC-MS data files. The combined HTML should be 'associated' with any one of the raw data files that has been used to generate the combined search. MSQuant can make use of the raw data file information in the peptide query to pick the correct file to open for peptide validation and quantitation purposes.

## 4. Questions or comments

Questions regarding this HOWTO may be directed to Shao-En Ong ([shaoen@bmb.sdu.dk](mailto:shaoen@bmb.sdu.dk))