

Exporting Scaffold Results as mzIdentML and MGF Files

Scaffold provides the ability to export peptide and protein identification information in the mzIdentML standard format for Proteomics data developed by the HUPO Proteomics Standards Initiative (Proteomics Informatics Standards group) and peaklists as MGF files.

To export Scaffold results, select the “mzIdentML...” option from the “Export” menu (Figure 1).

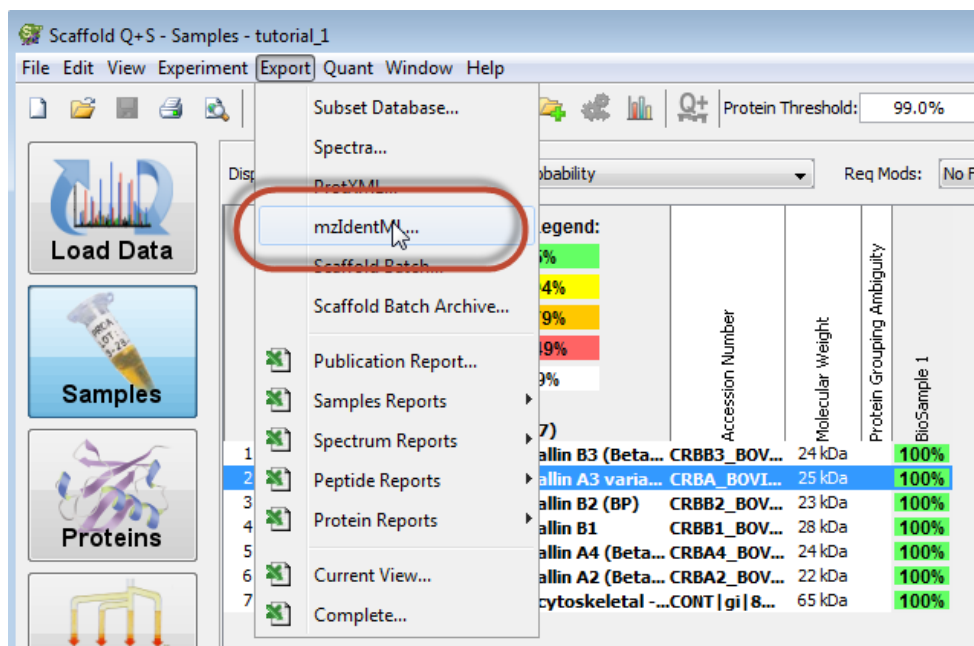


Figure 1 : Preparing to export mzIdentML/MGF files

This brings up the “Export mzIdentML” dialog box which provides basic options for creating mzIdentML exports optimized for different uses (Figure 2).

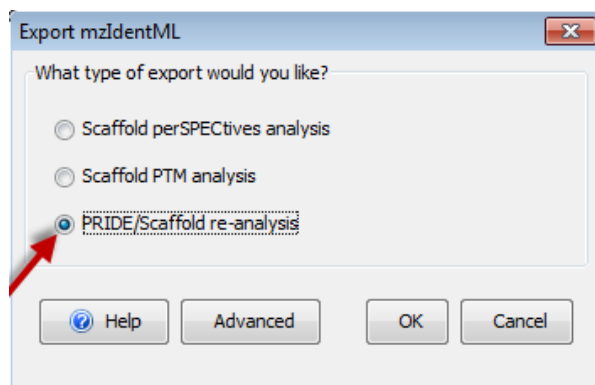


Figure 2: Export mzIdentML dialog

Each of these options configures a number of settings for the mzIdentML export. Clicking “OK” in this dialog launches the export process using the default settings for the selected option. These settings can be seen and adjusted by clicking the “Advanced” button. The default options for creating files for PRIDE submission are shown in Figure 3.

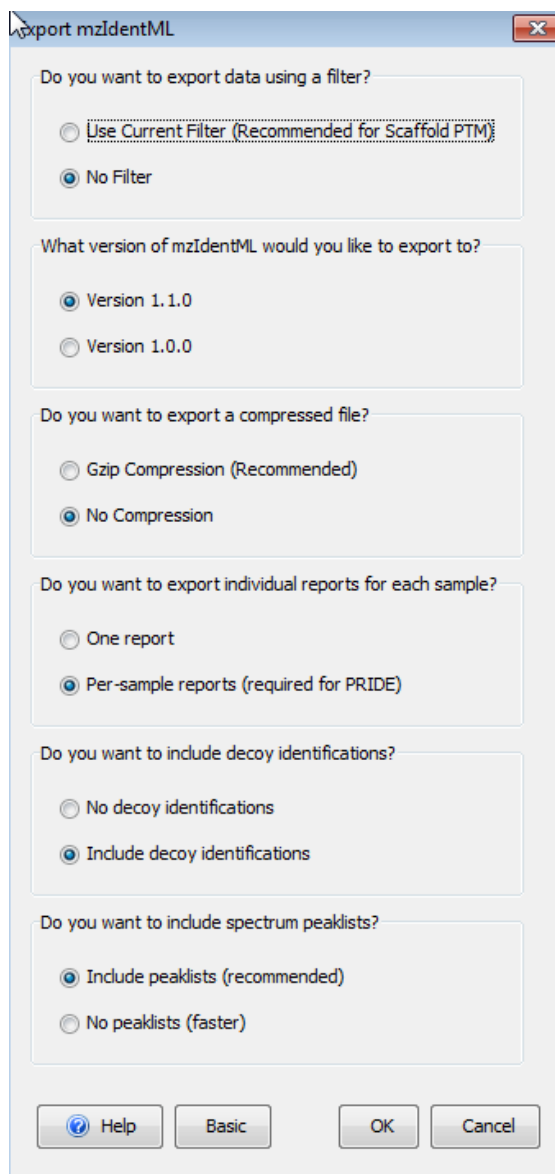


Figure 3: mzIdentML export options for PRIDE submission

For PRIDE submission, it is important to select the “Per-sample reports” option, as PRIDE expects each mzIdentML file to contain information about a single sample. Note that the “Include peaklists” option is selected by default. When this option is active, Scaffold automatically exports the spectrum peaklists as MGF files when it writes the mzIdentML files. Once all settings have been inspected and adjusted, the export may be launched by clicking “OK”

If there are any proteins for which Scaffold has not obtained the amino acid sequences, a warning dialog appears (Figure 4). If the only missing sequences are for decoy proteins, choosing “Export anyway” is recommended, but if many valid protein identifications show “?” in the molecular weight column of the Samples View, it may indicate that the FASTA database was not applied correctly during loading and this dialog provides an opportunity to remedy this situation before performing the export so that the mzIdentML file created will contain all of the information needed for subsequent processing.

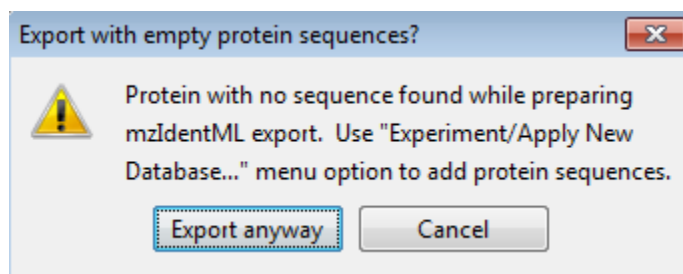


Figure 4: Warning of missing protein sequences

When the export begins, Scaffold brings up a File Browser to allow selection of the destination directory. Scaffold creates a folder in that directory with a name composed of the Scaffold experiment name and the date and time of export. Within this output folder, if the “Per-sample reports” option has been selected, Scaffold creates a directory for each BioSample and writes one .mzid and one .MGF for each MS-Sample within that BioSample. Otherwise, Scaffold writes a single .mzid file containing identification information for all samples and a .MGF for each MS-Sample in the entire experiment.