

# Scaffold Batch

Version 4.8

User's Manual

## Release information

The following release information applies to this version of the *Scaffold User's Manual*. This document is applicable for Scaffold, Release 4.8 or greater, and is current until replaced.

<i>Document Version Number</i>	<i>Scaffold_Batch_4.8</i>
<i>Document Status</i>	<i>Released</i>
<i>Document Release Date</i>	<i>June 2, 2017</i>

## Copyright

© 2017. Proteome Software, Inc., All rights reserved.

The information contained herein is proprietary and confidential and is the exclusive property of Proteome Software, Inc.. It may not be copied, disclosed, used, distributed, modified, or reproduced, in whole or in part, without the express written permission of Proteome Software, Inc.

## Limit of Liability

Proteome Software, Inc.. has used their best effort in preparing this guide. Proteome Software, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this guide and specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. Information in this document is subject to change without notice and does not represent a commitment on the part of Proteome Software, Inc. or any of its affiliates. The accuracy and completeness of the information contained herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every user.

The software described herein is furnished under a license agreement or a non-disclosure agreement. The software may be copied or used only in accordance with the terms of the agreement. It is against the law to copy the software on any medium except as specifically allowed in the license or the non-disclosure agreement.

## Trademarks

The name *Proteome Software*, the Proteome Software logo, *Scaffold*, *Scaffold Q+*, *Scaffold Q+S*, *Scaffold PTM* and *Scaffold perSPECTives* and the Scaffold, Scaffold Q+, Scaffold Q+S, Scaffold PTM and Scaffold perSPECTives logos are trademarks or registered trademarks of Proteome Software, Inc. All other products and company names mentioned herein may be trademarks or registered trademarks of their respective owners.

## Customer Support

Customer support is available to organizations that purchase *Scaffold*, *Scaffold Q+*, *Scaffold Q+S*, *Scaffold PTM* or *Scaffold perSPECTives* and that have an annual support agreement.

Contact Proteome Software at:

*Proteome Software, Inc.*

*1340 SW Bertha Blvd, Suite 10*

*Portland, OR 97219*

*1-800-944-6027 (Toll Free)*

*1-503-244-6027 (Direct)*

*1-928-244-6024 (Fax)*

[www.proteomesoftware.com](http://www.proteomesoftware.com)



<b>Chapter 1: Getting Started with Scaffold Batch .....</b>	<b>3</b>
Entering the License Key .....	6
Time based license key renewal .....	6
.....	8
<b>Chapter 2: Running Scaffold Batch .....</b>	<b>9</b>
Input and Output Files .....	9
Allocating Memory .....	10
Return codes: .....	10
Log files: .....	10
Monitoring progress: .....	10
Scaffold Batch and FASTA databases	11
Scaffold Batch and X! Tandem .....	12
<b>Chapter 3: Running Scaffold Batch through the Job Queue Wizard ..</b>	<b>15</b>
Create - specify a New Job .....	17
Running Jobs Through the Queue .....	26
<b>Appendix.....</b>	<b>29</b>
Sample XML Driver File .....	30
Annotated XML Driver (SCAFML) File Example	34



# Chapter 1

## Getting Started with Scaffold Batch

---

Scaffold is a software tool designed to help scientists identify and analyze proteins in biological samples. Using output files from MS/MS search engines, Scaffold validates, organizes, and interprets mass spectrometry data, allowing the user to more easily manage large amounts of data, compare samples, and search for protein modifications.

Scaffold Batch is a batch version of Scaffold. It can perform the same analyses as Scaffold does, but in a batch rather than in an interactive environment. The workflow for an experiment is defined in an XML driver file called a SCAFML, and Scaffold Batch is launched from a command line.

To simplify the process for users who are less experienced with command line applications, a Scaffold Batch wizard has been added to the regular Scaffold interface. Through this wizard, a user may create SCAFML files and run experiments in Batch mode.

Results from Scaffold Batch may be opened for further analysis in Scaffold or in the Scaffold Viewer. The Scaffold Viewer is a free, read-only version of Scaffold available Online for download. It facilitates the sharing of Scaffold analysis results among collaborators.

This chapter covers the following topics:

- [“Initial requirements” on page 4](#), which describes the minimum requirements for installing and running Scaffold.
- [“Installation” on page 4](#), which explains how to install Scaffold Batch 4.
- [“Allocating Memory” on page 6](#), which explains the memory requirements of the program.
- [Scaffold Viewer](#), which tells how Scaffold Batch users can share their results.

## Initial requirements

Before installing and running Scaffold the user needs to:

1. Be sure that the computer system where Scaffold will be installed and its network have access to directories containing:
  - Search engine output files for the samples that need to be analyzed
  - the FASTA database(s) used when those files were run.
2. Check the following document for general system requirement: [System\\_requirements.pdf](#). Check the following document for input files supported by the Scaffold Suite: [File\\_compatibility\\_matrix.pdf](#).
3. Check the following document for information on how to install the programs included in the Scaffold Suite: [installation\\_guide.pdf](#).
4. Check the following document for suggestions on how to set up searches with the most popular engines to optimize compatibility with Scaffold: [loading\\_search\\_engine\\_results\\_into\\_scaffold.pdf](#).
5. Have a license key to run Scaffold Batch, see [Installation](#).

## Installation

To install Scaffold Batch 4, begin by downloading the Scaffold 4 Program from:

<http://www.proteomesoftware.com/products/>

Scaffold Batch can be installed through the CLI or using the GUI (Desktop method):

### 1. Command Line Interface (CLI)

Running a command like the following will install Scaffold Batch:

**on Linux:**

```
bash Install_Scaffold_4XX_Linux_XX.sh > -c -q -dir /usr/local/scaffold_4.x.x
```

**on Windows:**

```
C:\users\username\Downloads>Install_Scaffold_4.8.0_windows-x64.exe -c -q -dir "C:\Program Files\Scaffold4"
```

Figure 1-1: Command options available for installer

Option	Function
-varfile [file]	Use a response file
-c	Run in console mode
-q	Run in unattended mode
-dir [directory]	In unattended mode, set the installation directory
-overwrite	In unattended mode, overwrite all files
-splash [title]	In unattended mode, show a progress window
-console	In unattended mode, open a console that shows the output
-manual	Manually select a Java runtime environment (not on Linux)
-Dname=value	Set system properties
-h	Show this help

## 2. Graphical User Interface (Desktop Method):

Click on the installer program. This will install both Scaffold and Scaffold Batch. For more information see

## Scaffold Tiered Licensing

The Scaffold suite of applications consists of the core Scaffold product, Scaffold Q+, and Scaffold Q+S, and Scaffold Batch. The core Scaffold product is the basis for all installations. The licensing key that Proteome Software provides determines whether the User has access only to the features and functions of the core Scaffold product, or the features and functions of Scaffold Batch, Scaffold Q+ or Scaffold Q+S.



*Users who purchase a license for Scaffold Batch also have access to all features of Scaffold. A Scaffold Q+ or Q+S license may also be combined with Scaffold Batch. Some, but not all, features and functions of Scaffold Q+ may be accessed in Batch mode. Quantitative data may be loaded and organized through Scaffold Batch and the resulting SF3 file may be opened in interactive mode for further analysis.*

Application	Description
Scaffold	Visualize and validate MS/MS proteomics experiments.
Scaffold Q+	Calculate and display relative protein expression levels in a sample determined by tandem mass spectrometry of iTRAQ- or TMT-labeled proteins.
Scaffold Q+S	Calculate and display relative protein expression levels in a sample determined by tandem mass spectrometry of stable isotopically-labeled (for example, SILAC) proteins.

Application	Description
Scaffold Batch	High-throughput automated processing by running Scaffold through a command line interface (CLI) or via the Scaffold Batch Wizard in Scaffold.

## Entering the License Key



*Scaffold Batch requires a special key and will not run even with a regular Scaffold evaluation or purchased key. To receive a Batch key, contact [support@proteomesoftware.com](mailto:support@proteomesoftware.com) or call (800) 944-6027.*

The license key can be installed either by starting Scaffold 4 in the interactive mode, or by running Scaffold Batch 4 at the command line. Follow the screen directions to insert the Scaffold Batch 4 key and complete installation. The Scaffold Batch 4 key will supercede all Scaffold 4 keys that had previously been installed on the computer. Since the Scaffold Batch 4 key unlocks Scaffold 4 as well, this will pose no conflict.

## Time based license key renewal

Time based license keys have time limits connected to the term of the user's support contract. When the contract expires, Scaffold still works but upgrades are not allowed until the support contract is renewed. To renew, fill out the **Key reset Request** page on the Proteome Software website and a sales representative will promptly respond providing further information.

## Allocating Memory

By default, Scaffold Batch 4 uses a maximum of 1024 MB of RAM. This is often not enough memory to handle large datasets. If more memory is available on the computer, more memory may be allocated to Scaffold Batch 4 by setting the java virtual memory options. To set the virtual memory available for Scaffold 4, edit the file

ScaffoldBatch4.vmoptions

This file is found in the Scaffold 4 installation directory.

The ScaffoldBatch4.vmoptions file contains a single line that looks like

`-Xmx1024m`

This tells java to allocate 1024 megabytes to Scaffold 4. To allocate four gigabytes, use “`-Xmx1g`” or “`-Xmx4096m`”. To allocate eight gigabytes, use “`-Xmx8g`”. To allocate 32 gigabytes, use “`-Xmx32g`”.

Scaffold 4 will not run reliably on less than 1024 MB. If the 32-bit version of the program is installed, no more than 2 GB may be allocated. It is recommended to install the 64-bit version if possible and to allocate as much memory to the program as possible while still

leaving some memory for the operating system and other applications.

## Scaffold Viewer

Scaffold Batch, like Scaffold produces output files in the SF3 format. These files can be opened by Scaffold or by the free Scaffold Viewer. The Viewer performs most of the functions included in a full Scaffold copy. However, it cannot load search results files and it can neither analyze data nor run X! Tandem.

With the Viewer, the User may share results with collaborators who can examine the data in the same ways as with Scaffold: by samples, proteins, peptides or spectra. Results can be filtered by protein probability, peptide probability and the number of matching peptides, by FDR values or by search engine scores. The Viewer allows a user to change the names of BioSamples and ,MS Samples, and can validate the peptide/spectrum matches. The Viewer user may also perform statistical analysis, annotate proteins with GO terms, etc.

Chapter 1  
Getting Started with Scaffold Batch

# Chapter 2

## Running Scaffold Batch

---

### Input and Output Files

The following file types are standard inputs and outputs from Scaffold Batch:

- **Inputs in Scaffold \*.sf3 (or \*.sfd)**

SEQUEST® = \*.dta and \*.out, \*.ms2 and \*.sqt, and \*.srf (BioWorks); Mascot® = \*.dat; X! Tandem® = \*.xml; Phenyx = \*.scaffold-pidres.xml; OMSSA = \*.omx; Waters® IdentityE = \*.xml (Note: These files must be created from our plugin in the new version of Waters PLGS); Proteome Discoverer = \*.msf For a complete list, see this document.

FASTA database files

- **Outputs – Scaffold \*.sf3 file**

protXML.xml file

mxIdentML (MZID) file

An XML driver file specifies these files and the parameters used to run Scaffold Batch. Example XML driver files and a specification for these files are given later.

To call Scaffold Batch from the command line or from a batch program:

```
scaffoldBatch4.exe [-f] [-q] [-keypath PATH] [xmlDriverFile]
```

Note that the Scaffold Batch key only has to be entered once, and future use of this command does not require this parameter. Also note that the “-f” option returns you to the prompt and requires no user input. This can be useful for checking program version, etc.

xmlDriverFile	This parameter is an XML file with commands which tell Scaffold Batch 4 which files to load. See the sections below “Sample XML Driver File” and “Annotated XML Driver File”.
---------------	---

Runtime options for the command line

-f	Force mode, don't request answers to questions from command line. This option will normally be used in all batch operations to keep Scaffold Batch 4 from stopping to wait for input.
-q	Quiet mode, write minimal information to standard output

## Chapter 2

### Running Scaffold Batch

-keypath PATH	Keypath is the path to the license key. Here “keypath” is the keyword and PATH is the fully qualified directory path to the license key file. This is only needed if the license key is not in the default Scaffold 4 directory. An instance where this might be useful is for grid systems.
---------------	--

The normal sample command line will look like this (on Windows):

```
scaffoldBatch4.exe -f driver1.xml
```

Scaffold Batch prints hundreds or thousands of lines to the system standard output. It may be useful to pipe this to a monitoring program. See [“Monitoring progress:” on page 10](#).

### Allocating Memory

By default, Scaffold Batch uses a minimum of 1024 MB of RAM. This is often not enough memory to handle large datasets. If more memory is available on the computer, allocating additional memory is advised. For details, see [“Allocating Memory” on page 6](#).

### Return codes:

Scaffold Batch returns a return code of 0 if successful and 1 if unsuccessful.

### Log files:

As Scaffold Batch executes it echoes the output log and the error log to the standard system console or standard out. It also writes this information to `output_log.txt` and `error_log.txt`. These files can be found in the directory where Scaffold 4 and Scaffold Batch are installed. On a Windows installation this is typically in the `C:\Program Files` directory.

The `output_log.txt` file is a record of the most recent execution of Scaffold 4 or Scaffold Batch. It includes numerous lines monitoring the progress of Scaffold 4. It is overwritten each time Scaffold 4 or Scaffold Batch starts.

The `error_log.txt` file is a cumulative record of all errors that Scaffold 4 or Scaffold Batch may have encountered. For debugging purposes, most of these errors are recorded with the Java system stack trace. Stack traces can make a small error seem rather fierce.

### Monitoring progress:

As Scaffold Batch 4 executes it prints its progress to the standard out file. Lines of the form

```
%9.47%
```

```
%9.59%
```

```
%9.7%
```

```
%9.8%
```

```
%9.91%
```

```
%10.01%
```

```
%10.12%
```

%10.24%

This output is provided so that the calling pipeline program can monitor this output stream and report the progress to the users. The numbers are the percent complete.

## Scaffold Batch and FASTA databases

When Scaffold Batch 4 runs, one of the parameters tells it which FASTA database to use to supply the protein sequence coverage and to use if Scaffold Batch 4 runs X! Tandem. Scaffold Batch 4 also needs to access an INDEX file for each FASTA database it uses. This INDEX may be present already or it may be created when Scaffold Batch 4 runs.

This index may be created in the interactive Scaffold 4 (Scaffold Desktop GUI) by using the menu option accessed by selecting Edit > Edit FASTA Databases... to get the Configure Database Parser or it may be created by specifying parse rules in the XML driver file for Scaffold Batch 4.

The behavior of Scaffold Batch 4 when dealing with FASTA databases and INDEX files is as follows:

- if the XML driver contains parse rules, they will be used both in reading the data and in applying the protein sequences. If there is already an INDEX file, it will be replaced.
- if the XML driver contains no parse rules but the database has been previously indexed (either in Batch or GUI) the existing index will be used in both phases.
- if no parse rules are included in the XML driver and the database has not been indexed already, Scaffold will make its best guess at a correct parse rule or use a generic rule to index the FASTA database.

If the FASTA file has non-standard header lines, you must specify parsing rules for the accession number and protein description. These parsing rules can be created and used in the interactive version of Scaffold 4.

The parsing rules can also be specified in the XML driver file for Scaffold Batch 4. The format of these parsing rules is that of regular expressions. An example is

```
<FastaDatabase
id="swissprot"
path="/Users/<user>/Documents/WorkFiles/control_sprot.fasta"
name="Generic"
databaseAccessionRegEx=">([ ]*)"
databaseDescriptionRegEx=">[ ]* (.*)"
/>
```

The use of back slashes to escape characters in the version of regular expression parsing is different than that used in the Mascot regular expression parser. Here are the parsing rules that the GUI version of Scaffold 4 provides as options.

name	databaseAccessionRegEx	databaseDescriptionRegEx
EST/NR (NCBI)	>(gi [0-9]*)	>[ <sup>^</sup> ]* (.*)
IPI (EBI)	>IPI:([ <sup>^</sup> \ .]*)	>[ <sup>^</sup> ]* Tax_Id=[0-9]* (.*)
Swiss-Prot (SIB/EBI)	>([ <sup>^</sup> ]*)	>[ <sup>^</sup> ]* \([ <sup>^</sup> ]*\) (.*)
UniProt/Swiss-Prot (UniProtKB)	>[ <sup>^</sup> ]*\([ <sup>^</sup> ]*\)	>[ <sup>^</sup> ]*\[ <sup>^</sup> ]* (.*)
UniRef/NREF (UniProt)	>UniRef100_([ <sup>^</sup> ]*)	>[ <sup>^</sup> ]* (.*)
Ensembl (EMBL/EBI)	>(ENS[ <sup>^</sup> ]*)	>[ <sup>^</sup> ]* (.*)
MSDB (Proteomics Group)	>([ <sup>^</sup> ]*)	>[ <sup>^</sup> ]* (.*)
Swiss-Prot Alt. Accession (SIB/EBI)	>[ <sup>^</sup> ]* \([ <sup>^</sup> ]*\)	>[ <sup>^</sup> ]*\[ <sup>^</sup> ]* (.*)
UniProt/Swiss-Prot Alt. Accession (UniProtKB)	>([ <sup>^</sup> \ ]*)	>[ <sup>^</sup> ]*\[ <sup>^</sup> ]* (.*)
Generic	>([ <sup>^</sup> ]*)	>[ <sup>^</sup> ]* (.*)

Sequest uses the parsing rules listed here as “Generic”. The parsing specified does NOT override the parsing in an existing index file.

## Scaffold Batch and X! Tandem

Some Scaffold users find it useful to search their Sequest or Mascot files a second time using X! Tandem. X! Tandem can be run within Scaffold Batch or externally as a standalone program.

There are two advantages to running X! Tandem within Scaffold Batch. The first is that it is convenient. The second is that since Scaffold controls X! Tandem it can easily merge the X! Tandem results back into the Sequest or Mascot results already loaded.

The main reason to run X! Tandem separately is that X! Tandem standalone can process many files simultaneously on a multi-threaded computer or computer cluster. A secondary reason is that this avoids some X! Tandem crashes that occur occasionally when X! Tandem is called from within Scaffold 4. You can download and install X! Tandem from [www.thegpm.org](http://www.thegpm.org). If you want to do a complex installation of X! Tandem you may want to contract for support with Beavis Informatics.

The main disadvantage of running X! Tandem standalone is that this makes Scaffold Batch 4 work much harder to match the results from X! Tandem with the corresponding results from Sequest or Mascot. This is a very RAM intensive part of Scaffold Batch 4 and may be a bottleneck in analyzing your data.

If you do decide to run X! Tandem standalone, there are certain parameters that you need to set in X! Tandem to make it compatible with Scaffold 4.

The main things you need are extra specifications (tags) in the default.xml/input.xml file. These XML files tell X! Tandem how to analyze the dataset. The extra tags we require

revolve around a) printing out publication information and b) reporting at least one match for every spectrum for statistical analysis. As such, the following tags are required:

1. This tag requires X! Tandem to include enough information so that we can line up spectra from different search engines:

`<note type="input" label="spectrum, dynamic range">1000.0</note>`

These tags force X! Tandem to search every spectrum and to not modify the spectrum:

1. `<note type="input" label="spectrum, use noise suppression">no</note>`
2. `<note type="input" label="spectrum, minimum parent m+h">0.0</note>`
3. `<note type="input" label="spectrum, minimum fragment mZ">0.0</note>`
4. `<note type="input" label="spectrum, minimum peaks">0</note>`

This tag requires X! Tandem to score everything it "identified":

1. `<note type="input" label="scoring, minimum ion count">0</note>`

These tags make X! Tandem put out the publication information and parameters we require:

1. `<note type="input" label="output, parameters">yes</note>`
2. `<note type="input" label="output, performance">yes</note>`

This tag requires X! Tandem to include the MS/MS spectra:

1. `<note type="input" label="output, spectra">yes</note>`

And this tag requires X! Tandem to provide results for every single spectrum (necessary for statistical analysis):

1. `<note type="input" label="output, results">all</note>`
2. `<note type="input" label="output, maximum valid expectation value">1000</note>`

For more information on the X! Tandem API see <http://www.thegpm.org/TANDEM/api/index.html>.

Chapter 2  
Running Scaffold Batch

# Chapter 3

## Running Scaffold Batch through the Job Queue Wizard

---

### Launching the Scaffold Batch Job Queue

The Scaffold GUI now offers an option for creating and running Scaffold Batch jobs. When the program is installed with a Scaffold Batch license, a special option is added to the File menu and an additional icon appears in the main tool bar of the program. Clicking on either of these options opens the Scaffold Batch Job Queue wizard.

The Scaffold Batch Job Queue allows a user to create Scaffold Batch driver (SCAFML) files through a simple graphical interface. Users less experienced with scripting and XML may find this easier than using the Command Line Interface.

Figure 3-1: Job Queue Option in Scaffold File Menu

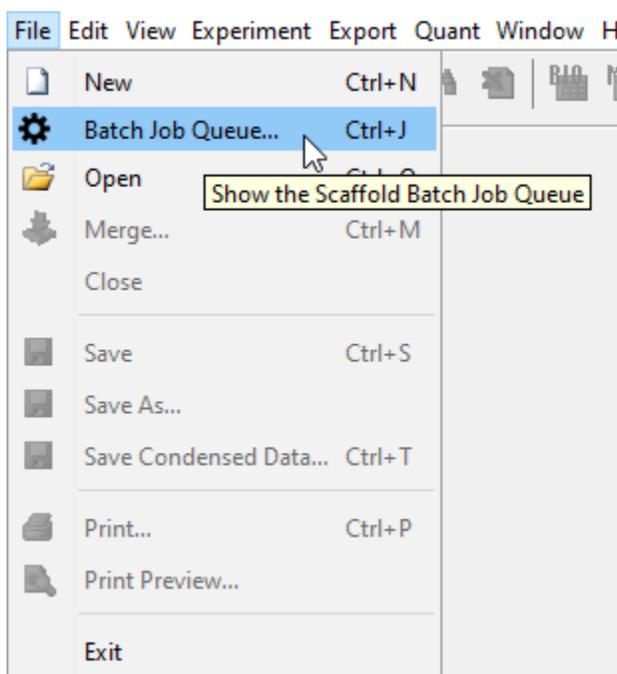


Figure 3-2: Job Queue Icon in the Scaffold Toolbar

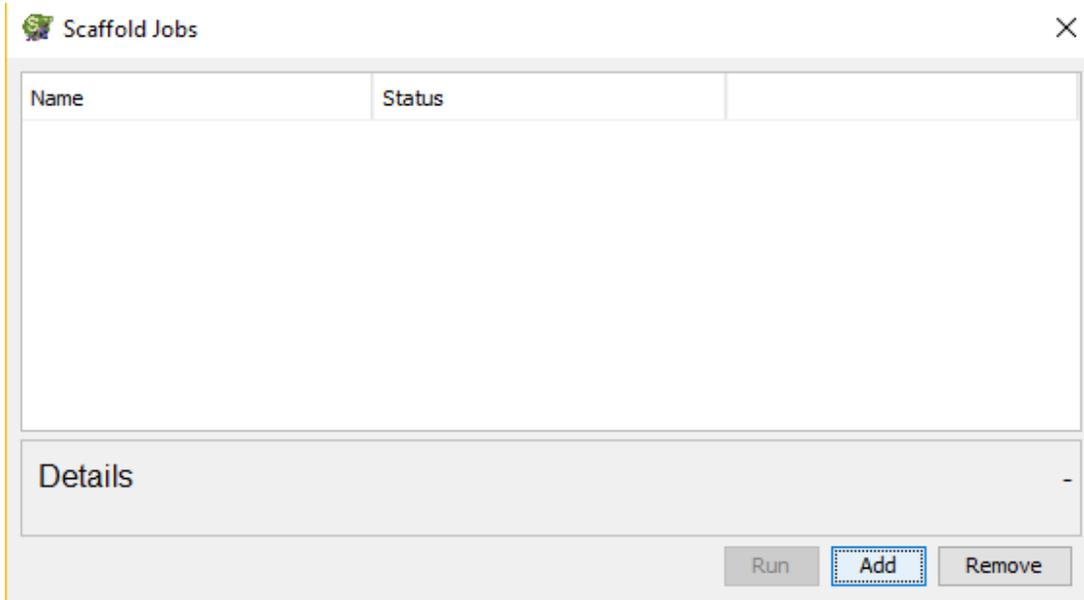


The Job Queue wizard supports most, but not all, Scaffold Batch functions. More complicated operations must be accomplished by editing the SCAFML files and running Scaffold Batch through the CLI.

## The Scaffold Batch Job Queue Wizard

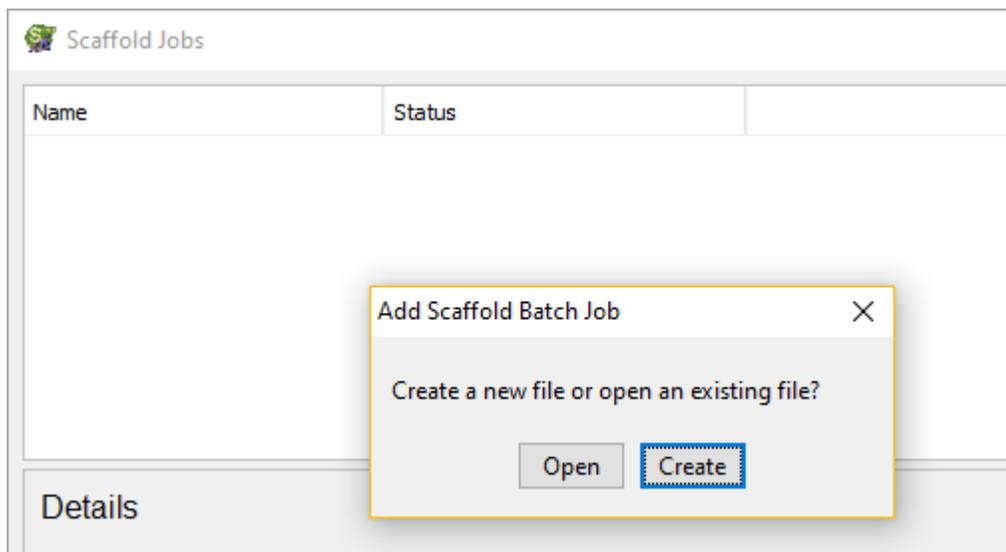
When the wizard is opened, it displays an empty Batch Job Queue.

Figure 3-3: Initial state of the Scaffold Batch Job Queue



Clicking Add presents the user with the option to Create a new job or Open an existing SCAFML for execution through the queue.

Figure 3-4: Adding a job to the queue



## Chapter 3

### Running Scaffold Batch through the Job Queue Wizard

#### Create - specify a New Job

When the Create button in the Add Scaffold Batch Job dialog is clicked, it presents a form through which the user can specify parameters for the Scaffold analysis to be performed in Scaffold Batch. This form is divided into a series of steps to help the user create the proper workflow for the new experiment.

The Scaffold Batch Job Queue Workflow Dialog

The screenshot shows a dialog box titled "Create Scaffold Job" with a close button (X) in the top right corner. The dialog contains a vertical list of nine steps, each with a numbered circle on the left and a small "+" or "-" sign on the right. Step 3, "Add Biological Samples", is highlighted with a pink background and has a dashed blue box around a button labeled "Add Biological Sample...". Step 4, "Use Existing Scaffold Files", is also highlighted in pink. Steps 1, 2, 5, 6, 7, 8, and 9 are in a light grey background. At the bottom of the dialog, there is a button labeled "Add biological samples or use an existing SF3", and two buttons labeled "Create" and "Cancel".

To the right of each step in the workflow dialog is a small “-” sign. Clicking on this sign expands the workflow step to allow access to all of its parameters. Default values are provided for many parameters, but steps which require attention before a SCAFML can be created are marked in pink.

A source of input data and at least one output file must be provided. If the user selects Biological Samples to add, the Use Existing Scaffold Files number will no longer appear in pink and vice-versa. It is possible to open an existing SF3 file and add new samples to it, however. Similarly, the user must either choose to save an SF3 file, create at least one export, or both.

### 1) Define Experiment

Figure 3-5: The Define Experiment Panel

The Define Experiment Panel allows the user to give a name and a description to the experiment, and provides drop-down menus for selection of the Protein Grouping method and the Scoring System (including Prefiltered Mode, in which Scaffold will not perform any scoring but will simply accept the search engine results, see the Scaffold User's Guide for more details.)

The Define Experiment Panel also contains the option to run X!Tandem during loading.

#### X! Tandem Analysis

The drop-down menu offers three choices:

- None- do not run X! Tandem during loading
- System Default - run the version of X! Tandem that is automatically installed with Scaffold
- Use Another Version of X! Tandem - allows the user to specify a different version of X! Tandem to be run. X! Tandem must have been already downloaded and installed on the computer on which Scaffold Batch is running. Selecting this option allows the user to browse to and select the X! Tandem executable file to be run during data loading.

**Use Only Identified Proteins** - when checked, Scaffold Batch creates a subset of the FASTA database containing only those proteins which were included in the loaded search results to confirm peptide identifications. If unchecked, X! Tandem searches the full FASTA database to try to identify additional proteins.

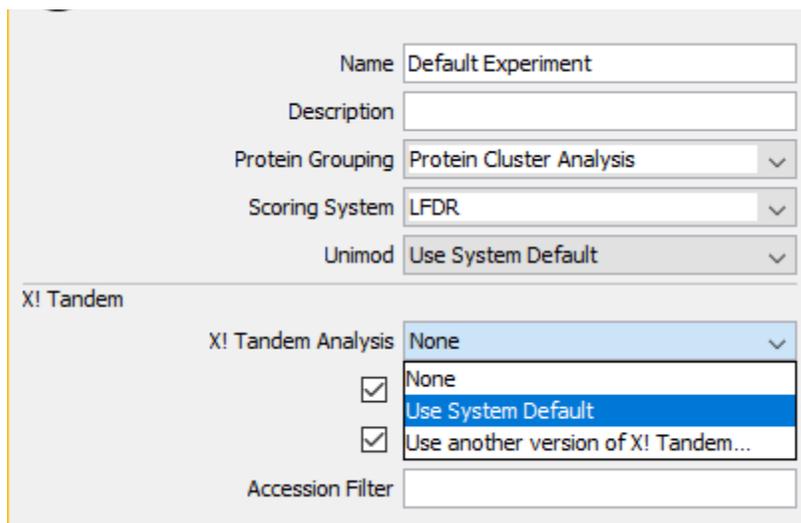
**Generate Decoy Proteins** - When checked, Scaffold Batch generates decoy proteins and adds them to the subset or FASTA database to be searched.

**Accession Filter** - allows the user to specify a text string that is used to filter the set of proteins to be included in the database that X! Tandem searches.

Figure 3-6: X! Tandem Options

At present, it is not to add modifications to the X! Tandem search through this dialog. The X! Tandem

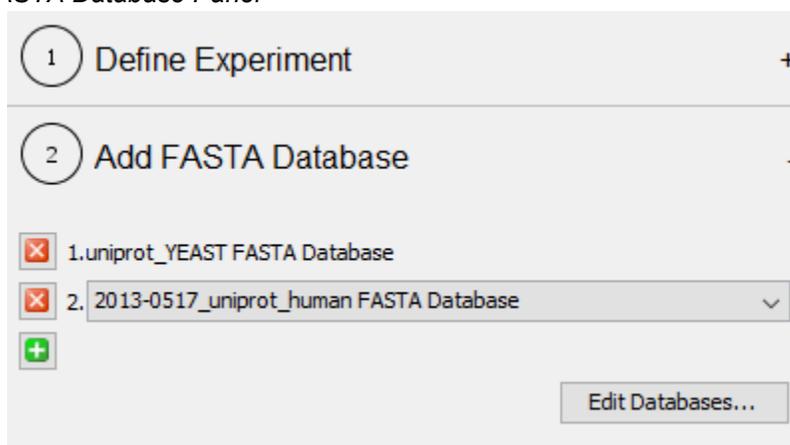
### Chapter 3 Running Scaffold Batch through the Job Queue Wizard



search will include all modifications from the original search(es) and the specific modifications that are always added by X! Tandem. The user may specify additional modifications by manually editing the SCAFML file, see Annotated SCAFML.

#### 2) Add FASTA Database

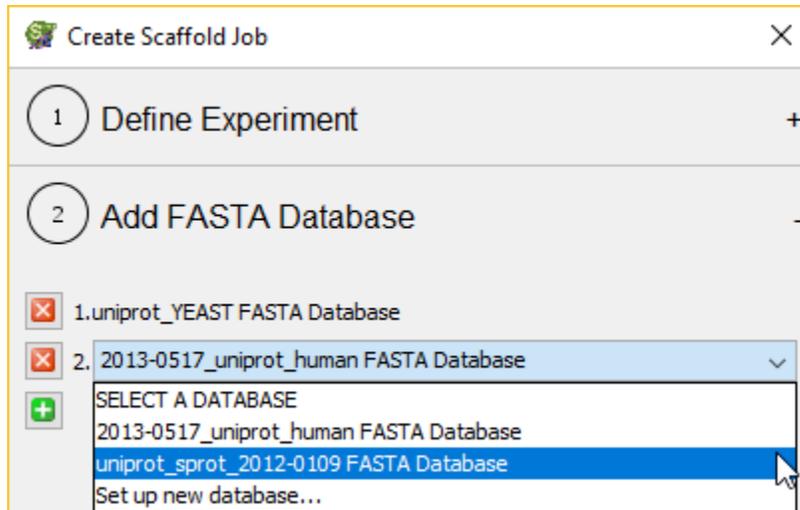
Figure 3-7: The FASTA Database Panel



The FASTA Database Panel displays the currently selected FASTA file (if any). A drop-down allows the selection of a different FASTA for the analysis. Using the “+” icon, the user can include additional FASTA files. The drop-down allows the user to select any already-parsed FASTA or set up new database.

**Edit Databases** - this button opens the full FASTA database dialog from the Scaffold interface.

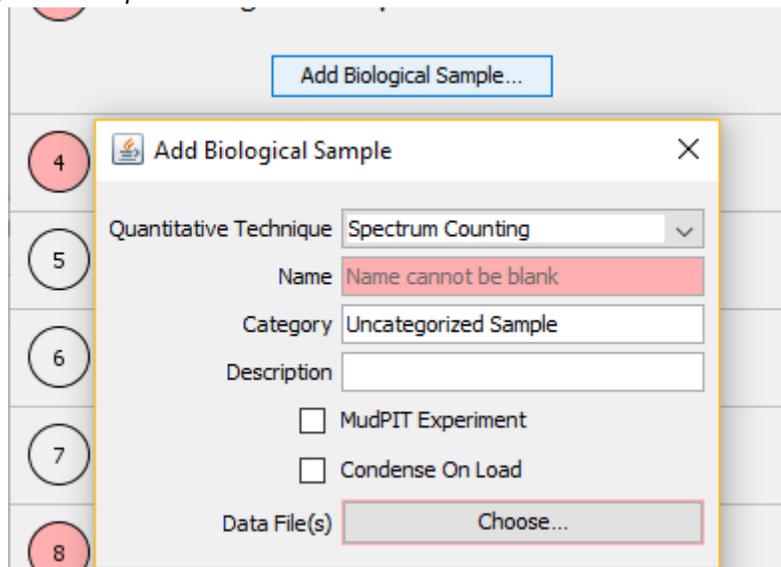
Figure 3-8: Select a FASTA



### 3) Add Biological Samples

This panel contains a button that opens a dialog for creating a BioSample in Scaffold.

Figure 3-9: Adding a BioSample



**Quantitative Technique** - select the quantitative method appropriate for the experiment:

- Spectrum Counting
- ITRAQ 4-Plex

### Chapter 3

#### Running Scaffold Batch through the Job Queue Wizard

- iTRAQ 8-Plex
- TMT 10-Plex
- TMT 6-Plex
- TMT 2-Plex
- Stable Isotope Labeling Multiplex
- TIC

**Name** - Required - the name of the BioSample

**Category** - Assigns the BioSample to a Category. Using the same Category name for multiple BioSamples groups them.

**Description** - a description of the sample.

**MudPIT** - Checkbox that determines whether the files loaded into this BioSample should be treated as a single MS Sample. This is useful when loading fractions.

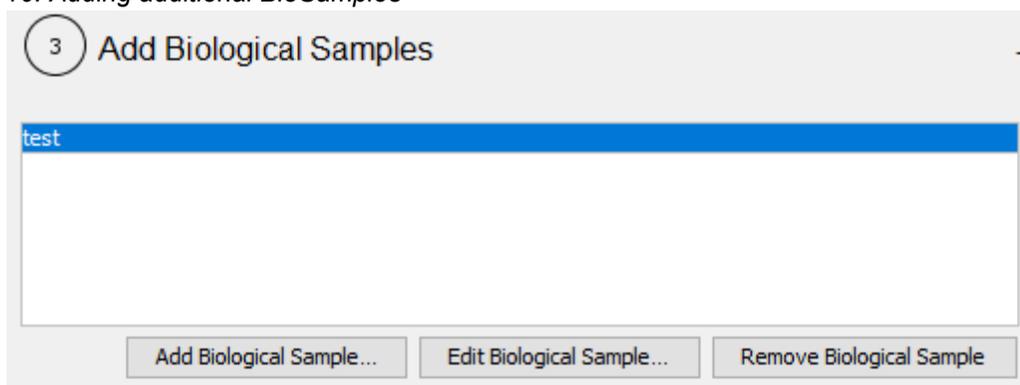
**Condense on Load** - If checked, low scoring spectra may be eliminated from the experiment.

**Data File** - Clicking on Choose opens a file browser from which the user may select files to be added to the BioSample. After all desired files are selected, the user should click Done.

**Add** - adds the Biological Sample to the Experiment.

After the Add button is clicked, the following dialog appears:

Figure 3-10: Adding additional BioSamples



The defined Biological Sample is displayed in the list, and the following options appear:

**Add Biological Sample** - Opens the Biological Sample dialog to create another BioSample.

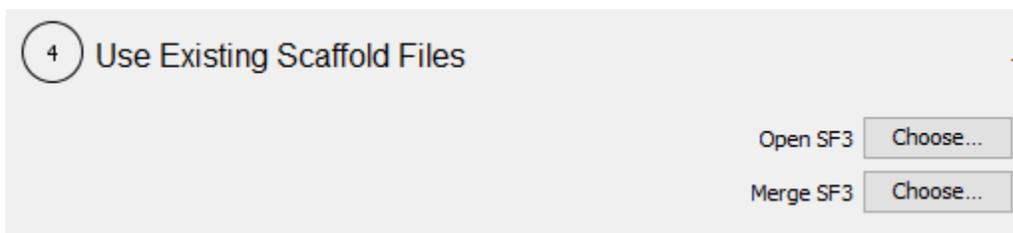
**Edit Biological Sample** - Reopens the Biological Sample dialog of the selected BioSample to allow editing.

**Remove Biological Sample** - Remove the selected BioSample from the list.

## 4) Use Existing Scaffold Files

This panel allows the user to open an existing SF3 file. There are three reasons for opening a file through Scaffold Batch:

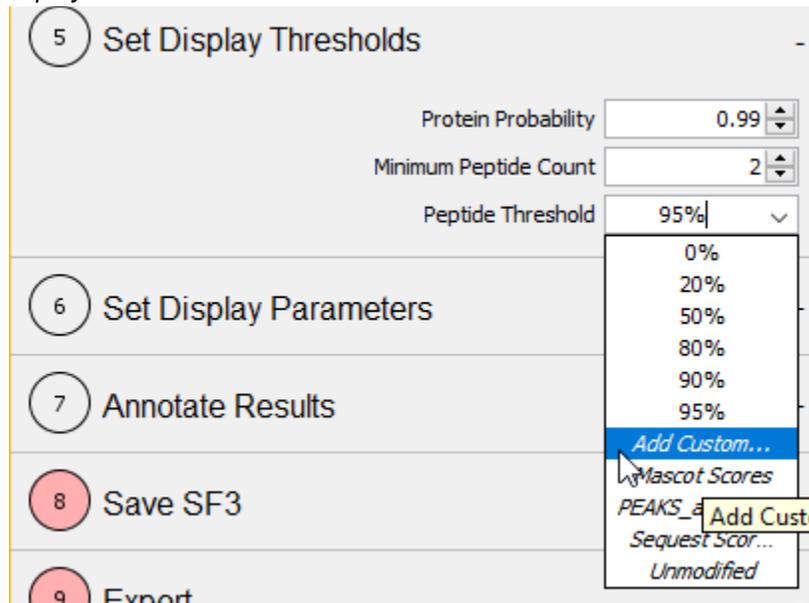
- To export reports - for example, a user might open a file and export an mzIdentML.
- To merge with another file - a control is supplied in this panel for choosing a file.
- To add data - a file may be opened and data specified through the Add Biological Sample panel may be added.



## 5) Set Display Thresholds

This panel allows the user to select the Display Thresholds which will be applied to the experiment. These thresholds will be used for exports, and will be shown in the SF3 file created by this experiment when it is opened.

Figure 3-11: The Set Display Thresholds Panel



Through this dialog, a user may select probability thresholds or Custom Peptide thresholds. It is also possible to open the Edit Peptide Threshold dialog by selecting Add Custom. At present, the Scaffold Batch Job Queue wizard does not support FDR level thresholds.

## 6) Set Display Parameters

This panel allows the user to select the parameters that will be used for display when any resulting SF3 file is opened. Some of the parameters also apply to exports.

Figure 3-12: The Set Display Parameters Panel

6 Set Display Parameters

Experiment Display Type: Protein Identification Probability

- Expand Clusters
- Show Similar Proteins
- Show Hidden Proteins
- Show Sample Notes
- View Biological Samples

Views

- Show Load Data View
- Show Samples View
- Show Proteins View
- Show Similarity View
- Show Quantify View
- Show Publish View
- Show Statistics View

Do not turn off ALL views.

## 7) Annotate Results

Through this panel, the user may apply GO terms, hide specific proteins, or star proteins of interest.

Figure 3-13: The Annotate Results Panel

**7 Annotate Results**

GOA Database

Annotate With NCBI

**Hide Proteins**

When Accession Number Matches

Regular Expression

Match Against Protein Name

Delimiter

**Star Proteins**

When Accession Number Matches

Regular Expression

Match Against Protein Name

Delimiter 

Color

**Publication**

Peak List Generator Name

Peak List Generator Version

Peak List Deisotoped

Peak List Charge States Calculated

Peak List Generator Annotation

Contains Common Contaminants

The Publication section allows the user to set the experimental variables which appear in the Publish View of the Scaffold interface. This is information that is needed for publication but which Scaffold is unable to extract from the search engine results.

## 8) Save SF3

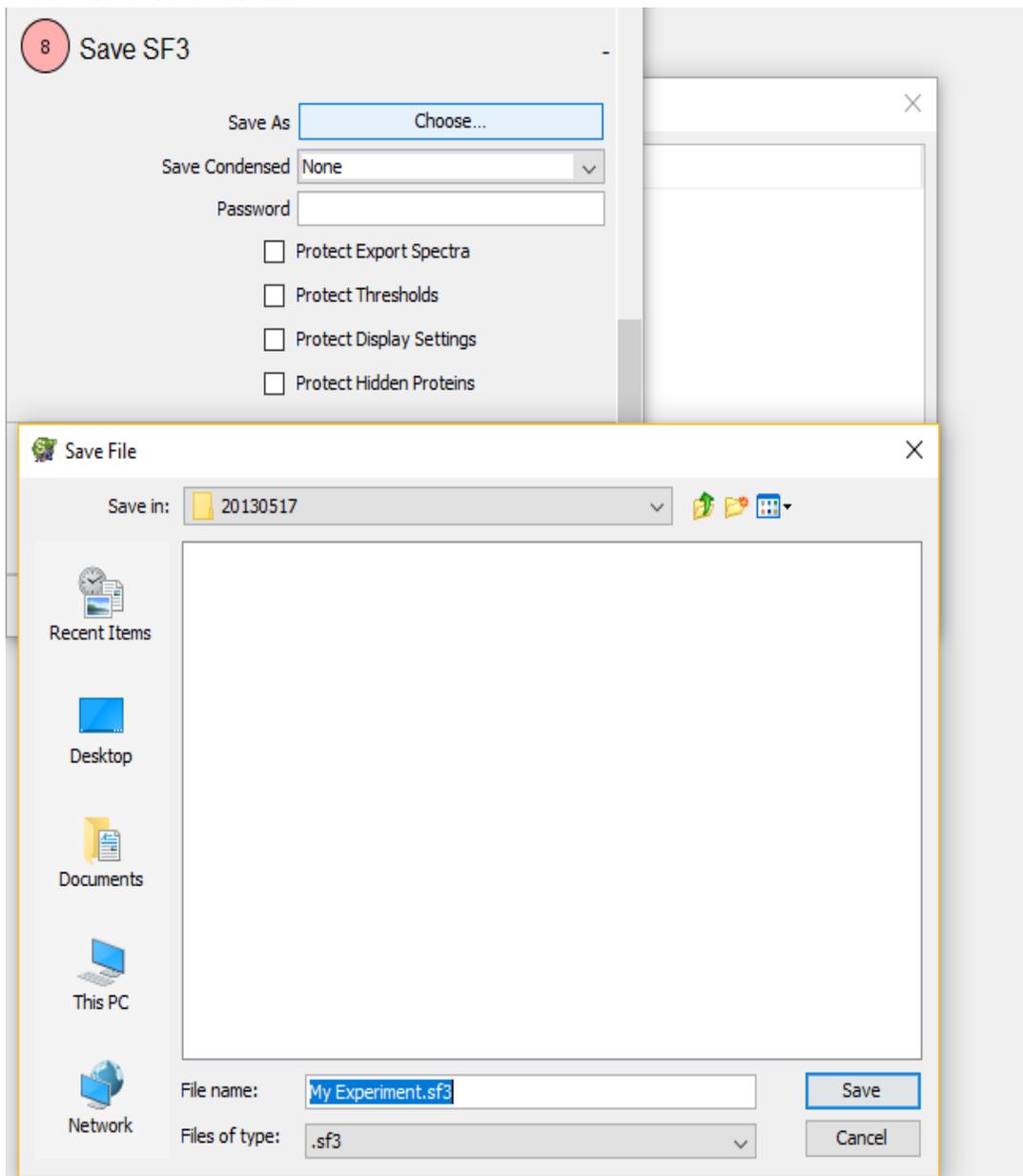
This panel provides the opportunity to specify a file name and location into which the Scaffold results should be saved at the conclusion of the analysis. In most cases, the user only needs to use the Choose button to open a file browser and specify the output file. This panel also allows the user to save the file with no spectra or with only a subset of the spectra to reduce the file size, lock the file with a password, or

## Chapter 3

### Running Scaffold Batch through the Job Queue Wizard

protect some of the settings in the file. This may be desirable if the file is to be shared with others whom the user does not want to have the ability to change the analysis.

Figure 3-14: The Save SF3 Panel

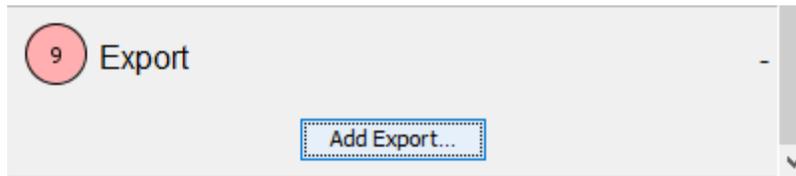


## 9) Add Export

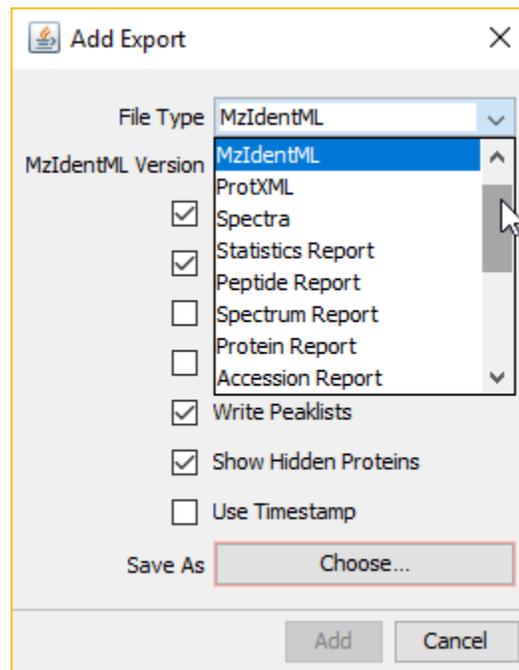
Results files such as mzIdentML or protXML and reports may be exported directly from a Scaffold Batch run.

Figure 3-15: The Report Panel

### Chapter 3 Running Scaffold Batch through the Job Queue Wizard



The Add Export button opens a dialog that allows the user to specify the parameters for any Scaffold report or export. Once the form has been completed, clicking Add adds the defined export to the export list. More than one export may be specified by repeatedly using the Add Export button.



## 10) Create

When all necessary information has been supplied through the Job Queue dialog, clicking the Create button opens a file browser. By default it opens in the location to which the experimental results are to be saved. The batch job must be named, and then saved as a SCAFML file.

## Running Jobs Through the Queue

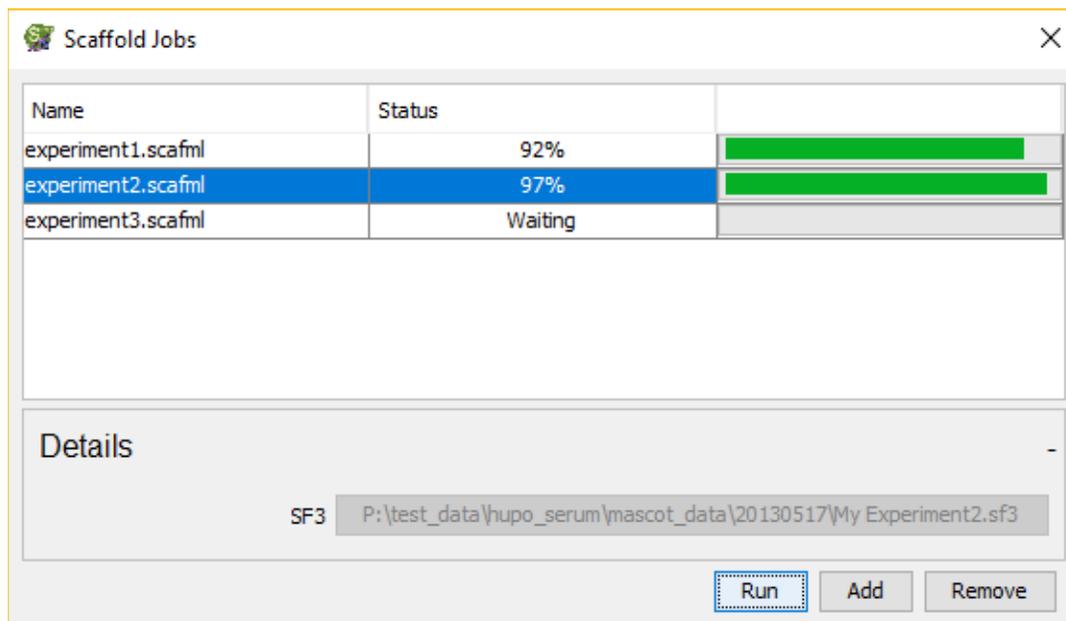
When a SCAFML is created through the Scaffold Batch Job Queue wizard, it is added to the Batch Job Queue, which is displayed to the user.

*Figure 3-16: Running Jobs in the Batch Job Queue*

Through this dialog, the user can:

## Chapter 3

### Running Scaffold Batch through the Job Queue Wizard



- **Run** - run the Scaffold Batch process. Multiple Scaffold processes may be run at the same time with Batch.
- **Add** - add an existing SCAFML file to the queue to be run. Note that more complicated SCAFML files may not be able to be loaded into the queue and may need to be run from the Command Line.
- **Remove** - remove a job from the queue.

When a job is placed on the queue, its Status is displayed as Waiting. While it is running, the Status column shows the percentage of the process that has completed. The green bar at the right tracks the progress.

## Details Pane

The Details Pane may be expanded or collapsed by clicking on the "+" or "-" sign at right.

The user may select a job in the queue by clicking on its row, and the output files specified in the selected SCAFML will be displayed in the Details pane. When the job is complete, the output files become clickable, and clicking on a file name opens the file. For SF3 files, clicking will open Scaffold as a Viewer, since only one instance of interactive Scaffold may be run at a time.

## Error Reporting

If a Scaffold Batch run produces an error, the condition that caused the error is displayed in the Details Pane, and is highlighted in pink.

Chapter 3  
Running Scaffold Batch through the Job Queue Wizard

# Appendix

---

- [Appendix A. Sample XML Driver File](#)
- [Appendix B. Annotated XML Driver \(SCAFML\) File Example](#)

## Appendix A. Sample XML Driver File

Scaffold Batch requires an XML driver file to define its operational parameters. This documentation provides two sample XML driver files and an annotated list of the tags in the XML driver file.

Example of a minimal Scaffold Batch XML file:

```
<Scaffold>
  <Experiment name="sb_test_1">
    <FastaDatabase id="swissprot" path="C:\sprot.fasta"/>
    <BiologicalSample database="swissprot" >
      <InputFile>C:\control_01.dat</InputFile>
    </BiologicalSample>
    <BiologicalSample database="swissprot">
      <InputFile>C:\control_02.dat</InputFile>
    </BiologicalSample>
    <DisplayThresholds name="DTs" id="thresh"
proteinProbability="0.95"
      minimumPeptideCount="2" peptideProbability="0.95"/>
    <Export type="sfd" thresholds="thresh"
      path="C:\sb_test_1.sfd"/>
    <Export type="protxml" thresholds="thresh"
      path="C:\sb_test_1_protxml.xml"/>
  </Experiment>
</Scaffold>
```

Example of a more complete Scaffold XML driver file:

```
<Scaffold>
  <Experiment name="Demo"
    description="An Example Experiment"
    load="Demo_existing.sfd"
    peakListGeneratorName="Bioworks"
    peakListGeneratorVersion="3.2"
    peakListDeisotoped="false"
```

```

    showStatisticsPane="false"
    password="foo"
    protectThresholds="true"
    protectDisplaySettings="true"
    protectExportSpectra="true"
    analyzewithTandem="true"
    condenseDataWhileLoading="true">
<FastaDatabase
  id="swissprot"
  name="swissprot"
  databaseAccessionRegEx=">([ ]*)"
  databaseDescriptionRegEx=">[ ]* (.*)"
  path="../Sprot_human.fasta"/>
<Modification unimodName="HNE"/>
<Modification
  fullName="testMod"
  referenceName="a test mod"
  monoMass="42"
  averageMass="42"
  minoAcidsModified="A"/>
<BiologicalSample
  analyzeAsMudpit="true"
  database="swissprot"
  name="Demo_01"
  category="Category 1">
<InputFile>Demo_01.dat</InputFile>
<InputFile>Demo_01.tar.gz</InputFile>
</BiologicalSample>
<BiologicalSample
  analyzeAsMudpit="true"
  database="swissprot"
  name="Demo_02">
<InputFile>Demo_02.dat</InputFile>
<InputFile>Demo_02.mgf.xml</InputFile>
< InputFile>Demo_02.tar.gz</InputFile>
</BiologicalSample>

```

```

<BiologicalSample
  analyzeAsMudpit="true"
  database="swissprot"
  name="Demo_03">
<InputFile>Demo_03.dat</InputFile>
<InputFile>Demo_03.mgf.xml</InputFile>
<InputFile>Demo_03.tar.gz</InputFile>
</BiologicalSample>
<DisplayThresholds
  name="Some Thresholds"
  id="thresh"
  proteinProbability="0.95"
  minimumPeptideCount="2"
  peptideProbability="0.95"
  minimumNTT="1"
  useCharge="true,true,true"/>
<Export
  type="sfd"
  thresholds="thresh"
  saveOnlyIdentifiedSpectra="true"
  saveNoSpectra="true"
  path="."/>
<Export
  type="protxml"
  thresholds="thresh"
  path="Demo_foo.protxml"/>
</Experiment>
</Scaffold>

```

Example of a file merge xml driver file:

```

</Scaffold>
<Experiment
  name="control_experiment"
  description="An Example Experiment"

```

```
    file="/Scaffold/Scaffold Analysis/11-19-07/2_out.sfd"  
    merge="/Scaffold/Scaffold Analysis/11-19-07/1_out.sfd">  
  </Experiment>  
</Scaffold>
```

(Merges 11-19-07/2\_out.sfd file with 11-19-07/1\_out.sfd)

For a fully annotated Scaffold Batch XML driver file (SCAFML example file) see [“Annotated XML Driver \(SCAFML\) File Example”](#) on page 34.

# Appendix B. Annotated XML Driver (SCAFML) File Example

*See Annotated SCAFML File*



# Index

## A

Appendix ..... 29

## C

Copyright ..... 2

## L

License key registration

    License key renewal ..... 6

licensing for Scaffold..... 5

## R

Release Information..... 2

Renewing time based license key.. 6

## S

Scaffold

    tiered licensing for ..... 5

## T

Time based license key renewal.... 6