PEAKS® CMD

PROTEOMIC COMMAND LINE SOLUTION

Microsoft Windows User Guide

November, 2015

Bioinformatics Solutions Inc.
www.bioinfor.com

1. Introduction

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) based proteomics has become a routine approach in life science since the past decade. In a typical proteomics lab, vast amounts of experimental data are generated daily requiring batch analysis without manual intervention. Further, a growing number of institutions utilize analytical pipelines to perform fundamental proteomic identification tasks. To address each of these cases PEAKS CMD offers a command line-based solution for protein/peptide identification.

The following describes how to set up and configure PEAKS CMD on a Microsoft Windows computer ensuring prompt sample analysis either independently or as part of a customized pipeline.

2. Configuration

To meet the demands of analyzing large data sets, PEAKS CMD is designed for 64 bit Windows Operating Systems (XP, Vista, and 7). The amount of disk space required depends on the size of the user datasets. The two main factors affecting PEAKS performance are CPU and RAM.

- A recommended configuration can be an Intel Core processor, 8GB RAM.
- An ideal configuration may be closer to Intel Core i7 or Xeon processors, 16GB RAM or more (or 2GB per core).

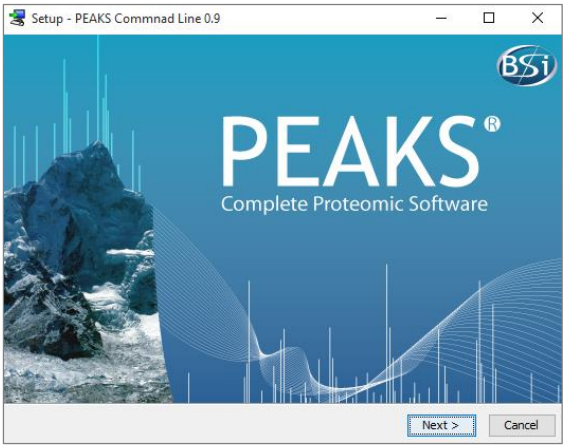
3. Installation

***Step 0*:** Download the PEAKS CMD setup file to the intended computational resource (pipeline core or computer). In the event this file cannot be found, please contact sales@bioinfor.com. They can also assist with registration key issues.

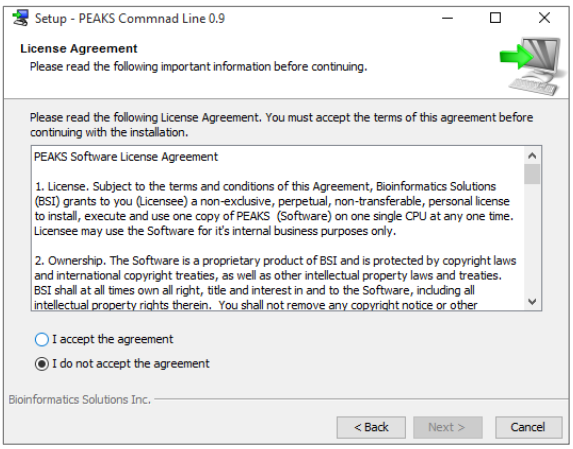
***Step 1*:** Locate to the PEAKS CMD software package directory and double-click the setup file to launch the installation. The setup file can be easily recognized by its icon  .

NOTE: If PEAKS CMD is intended to be installed on a Microsoft Windows Server, right-click the setup file and select “Run it as administrator”.

***Step 2*:** Click the “Next” button to start the installation process.

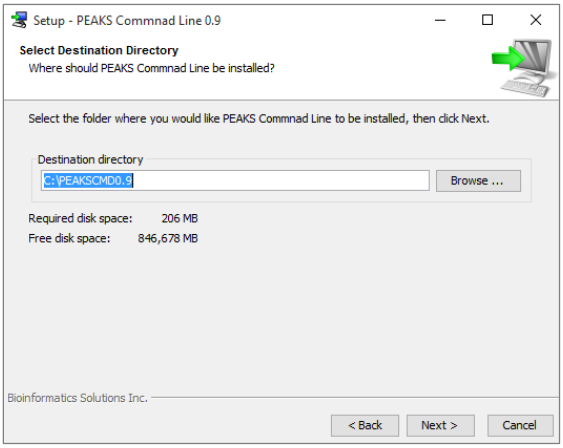


***Step 3*:** Check the radio button “I accept the agreement” and click “Next”.

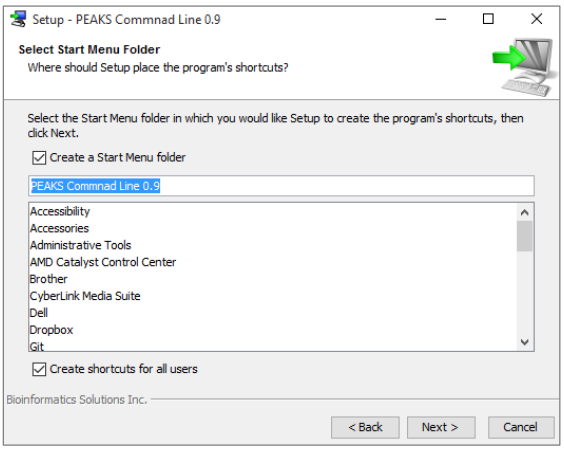


***Step 4*:** Select the desired installation directory for PEAKS CMD using the “Browse …” button and click “Next >”.

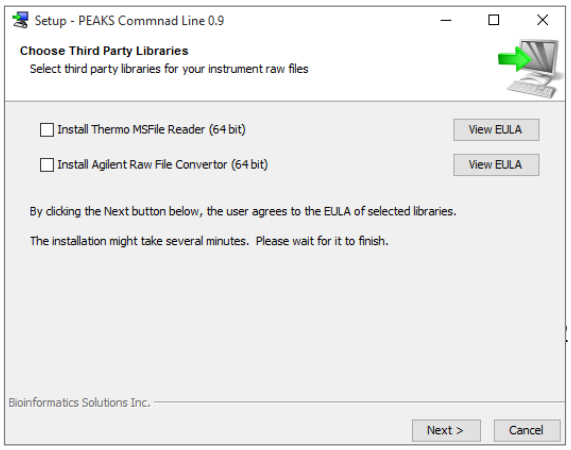
NOTE: From this step users can simply click the “< Back” button to navigate to the previous step for possible changes.



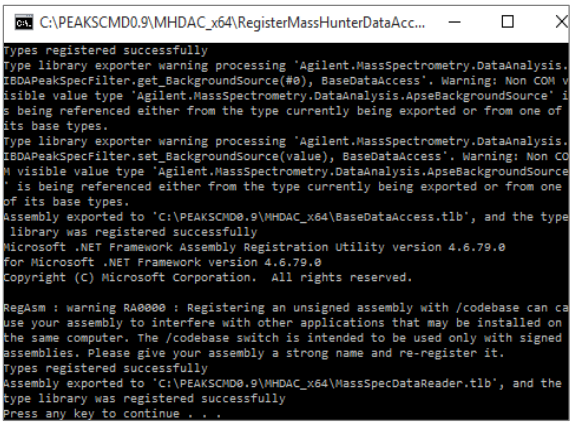
***Step 5*:** Create the Start Menu Folder, and click “Next >”.



***Step 6*:** Select the boxes of any required third party libraries, such as Thermo or Agilent, for direct loading instrument raw files. Click “Next >”. This will prompt the installation of third party software onto this resource.

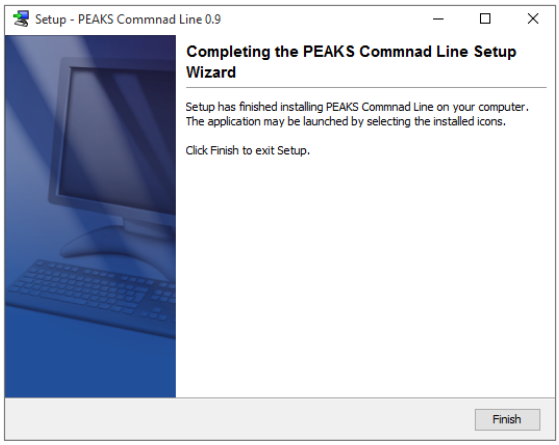


NOTE: A DOS window may appear requesting "press any key to continue", press any key to process to the next step.



NOTE: If third party software installations are not required, click "Next >" directly.

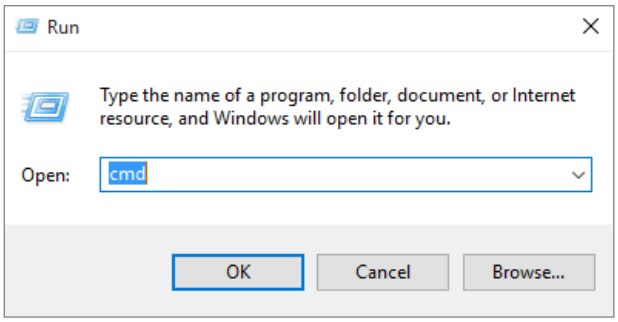
***Step 7*:** The following dialog indicates the successful installation of PEAKS CMD. Click "Finish" to complete the installation.

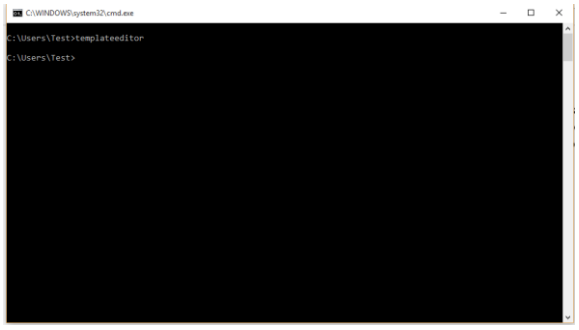


4. Registering PEAKS CMD

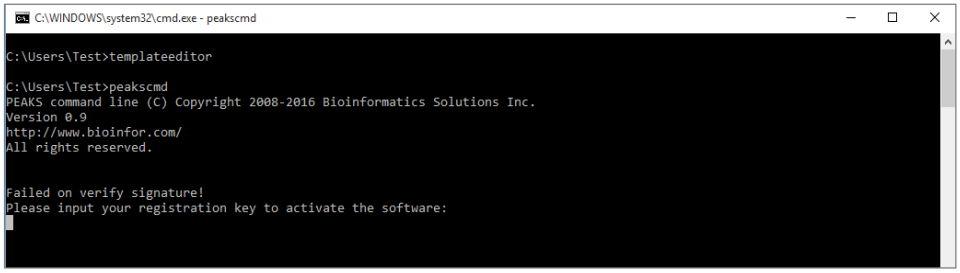
Microsoft Windows command line must be launched to run PEAKS CMD. To launch a command line, click Windows "Start" button or " + S", type "cmd" in the Search or Run line and press Enter. Typically, a Windows command line will start to run as shown below:

NOTE: If PEAKS CMD is installed on a Microsoft Windows Server, you need to run Windows command line as administrator.



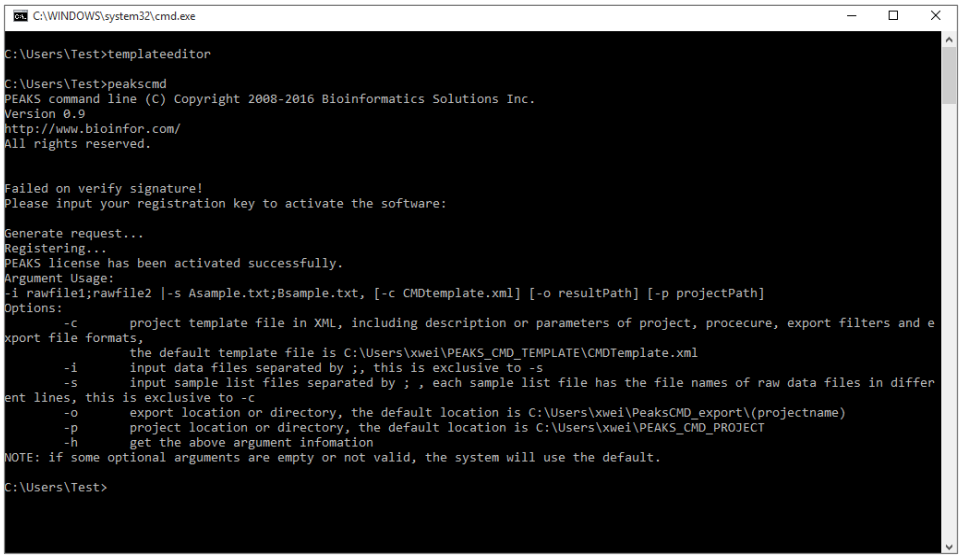


Type “peakscmd” in the Windows command line to run PEAKS CMD. The first time PEAKS CMD is run, it will ask for a registration key to activate the software.



Type or paste in the registration key acquired from a BSI sales representative.

Usage information will be shown if the registration is successful.

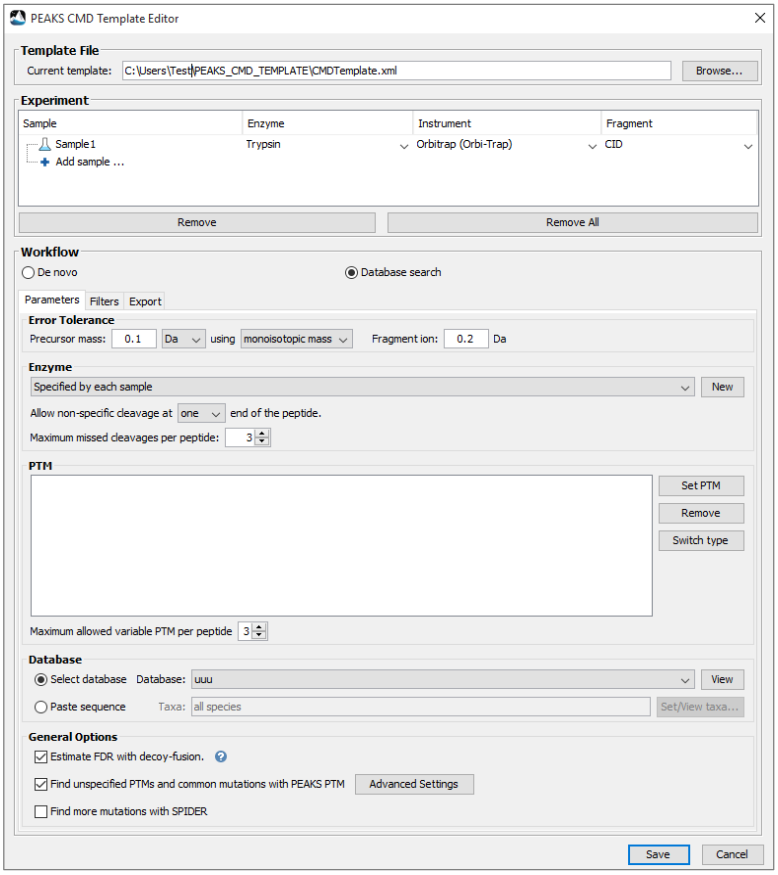


5. Using PEAKS CMD

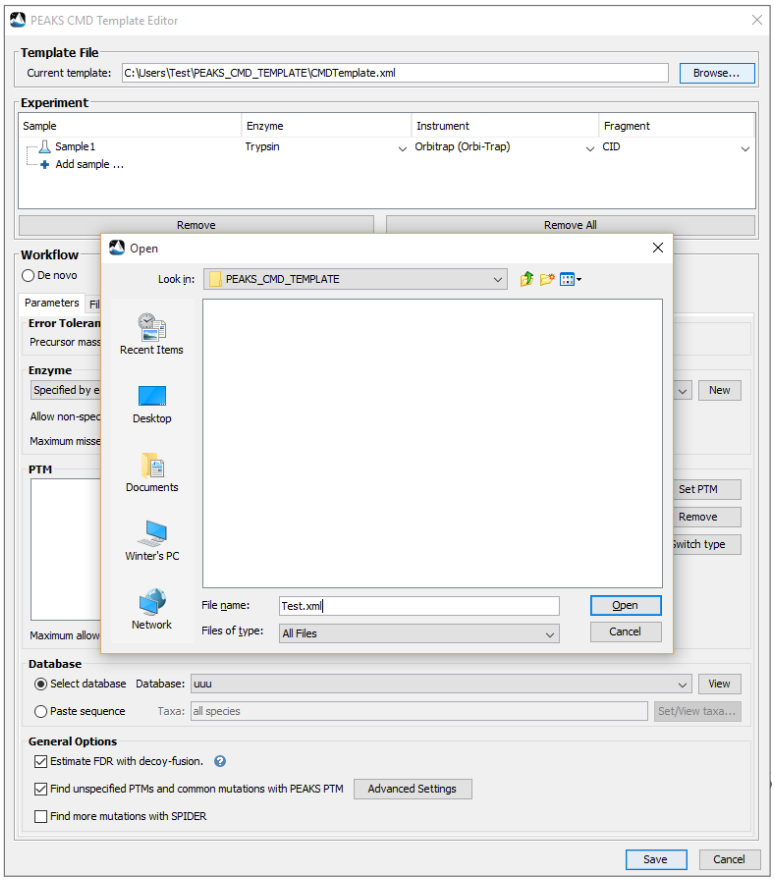
Performing protein/peptide identification using PEAKS CMD requires two steps: 1) generating a template file and 2) running PEAKS CMD with a template file.

5.1 Generating a Template File

The PEAKS CMD Workflow, including configurations of MS file loading, search types, parameters, protein/peptide filtration, and result exporting, is customized by users and stored in a template file. To run PEAKS CMD tasks, users must first create a new (or edit an existing) template file. A GUI-based editor, TemplateEditor, is provided to create or edit template files. Users can launch the editor by typing "TemplateEditor" in Windows command line and press Enter. The editor will appear as presented right:



***Step 1*:** Create or edit the template file in the "Template File" section. The path shown in the "Current Template" text field is the default path to save a template file. Users can also save template files in a preferred directory and with different files name by clicking "Browse…".

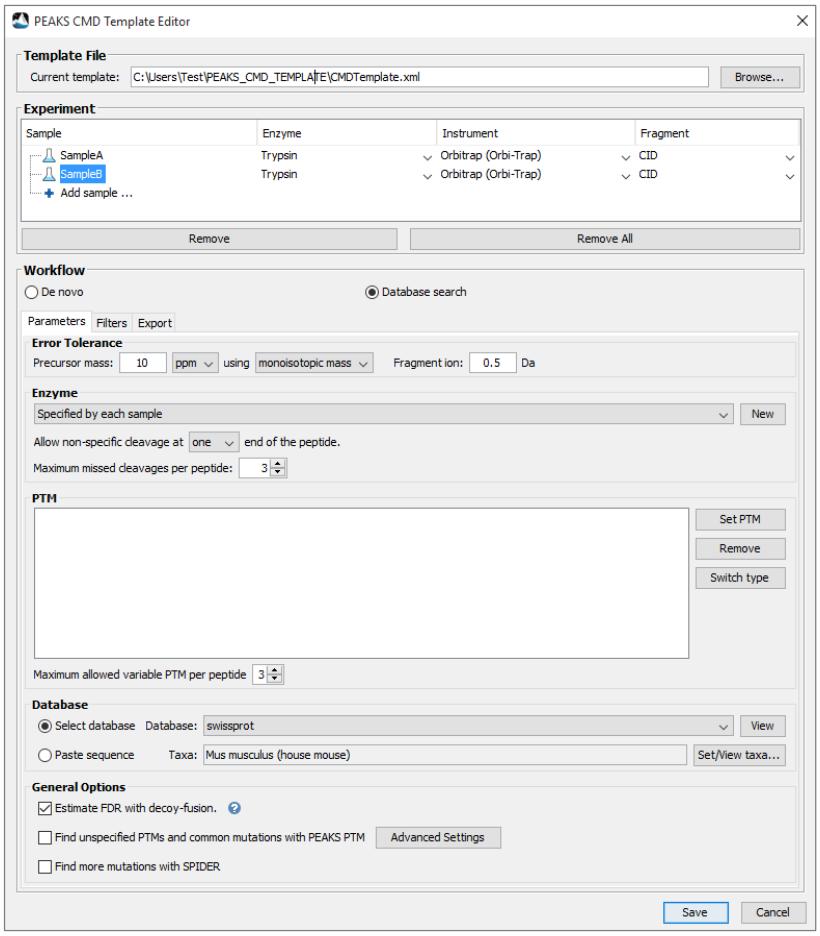


***Step 2*:** Add or remove samples in the "Experiment" section. PEAKS CMD supports analyzing an experiment consisting of multiple samples; add a sample by clicking the "Add sample" or remove selected samples by clicking the "Remove" button. Clicking the "Remove All" button removes all the samples listed in the experiment.

NOTE: Adding sample files in this step is not necessary. Users only need to set the sample numbers and names for the experiment.

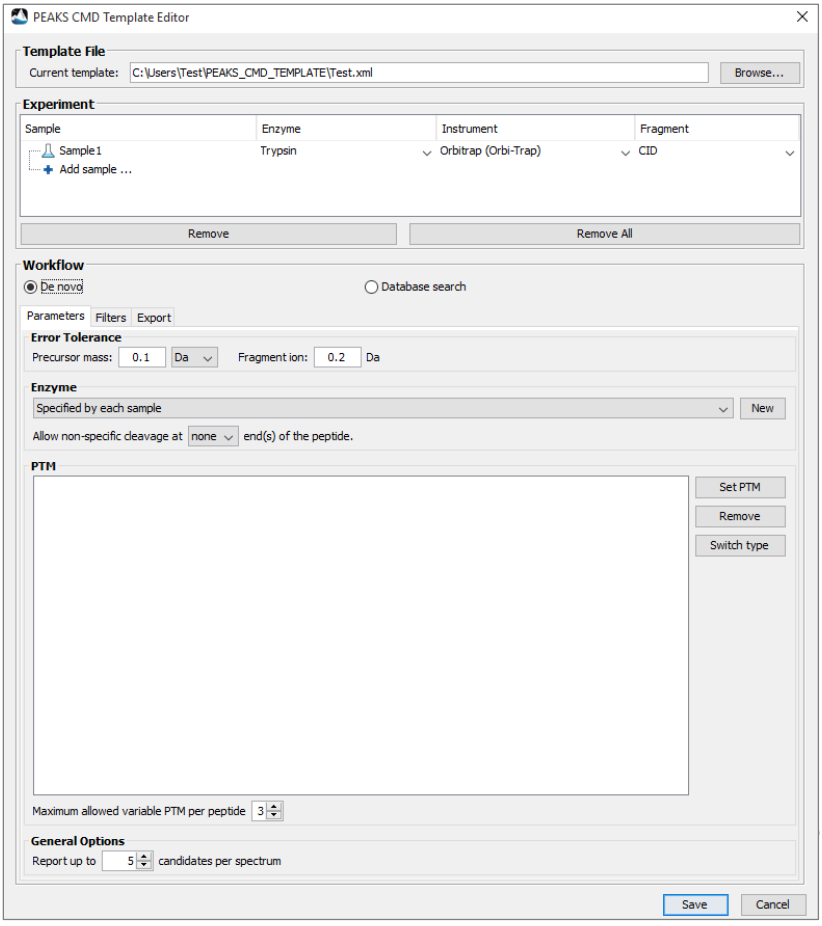
***Step 3*:** Select a task for PEAKS CMD. Currently two types of PEAKS CMD tasks are provided in the "Workflow" section: 1) "De novo" for *de novo* sequencing, and 2) "Database Search" for PEAKS database search.

***Step 4*:** Configure PEAKS CMD task parameters. Typically, *de novo* sequencing and database searching require slightly different parameter configurations. The following steps demonstrate how to configure the parameters required in *de novo* sequencing and database searching, respectively.

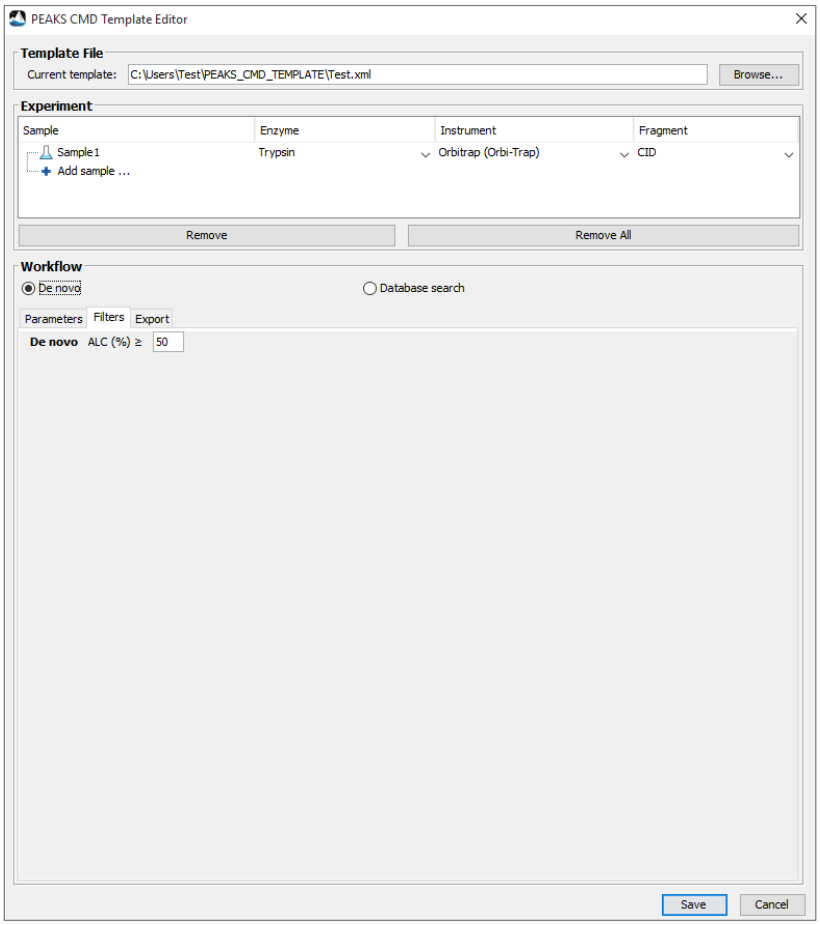


***Option 4.1*:** Check the "De novo" radio button.

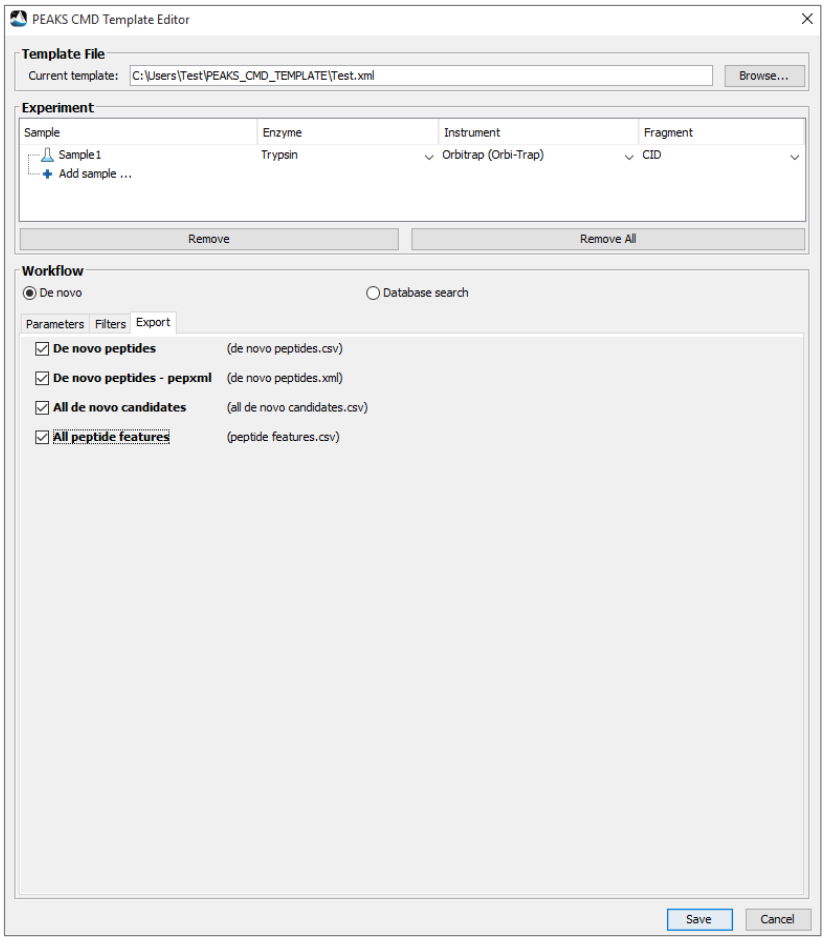
***Step 4.1.1*:** Configure the *de novo* sequencing parameters under the "Parameters" tab.



***Step 4.1.2*:** Set the *de novo* sequencing result filter under the "Filters" tab.

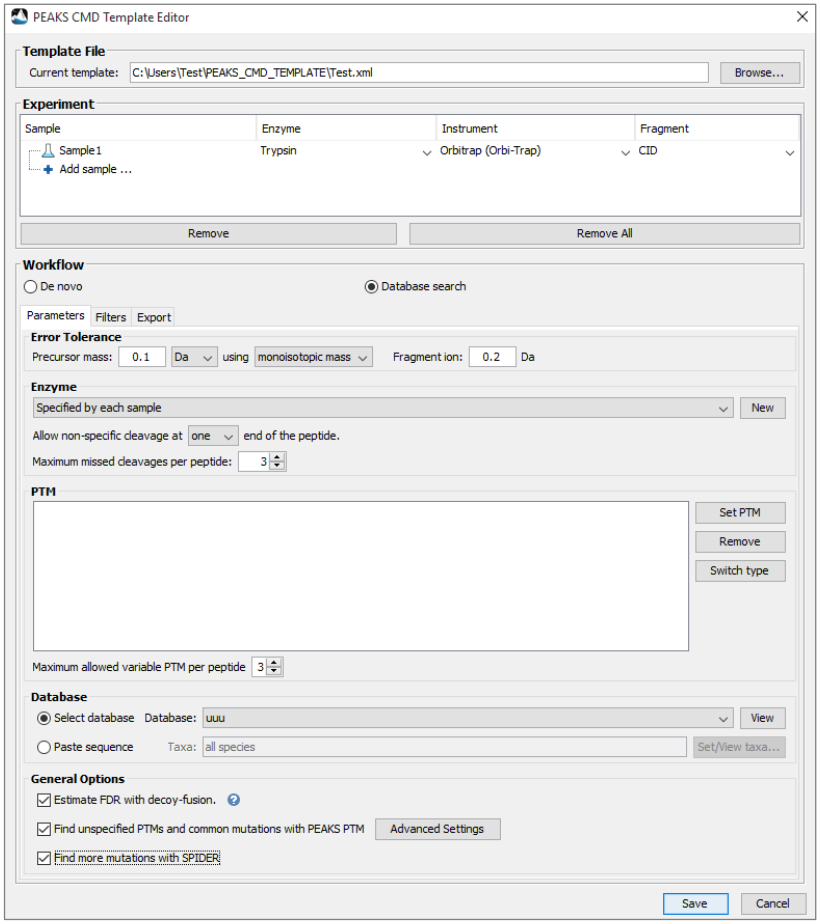


***Step 4.1.3*:** Select the files required for export under the “Export” tab.



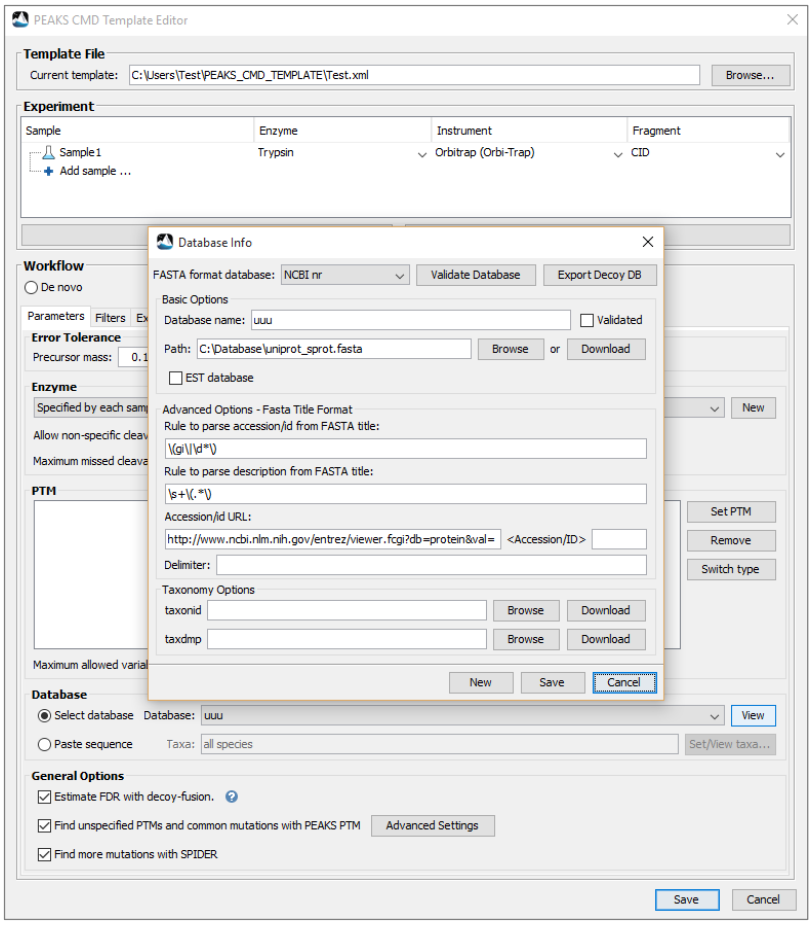
***Option 4.2*:** Check the "Database search" radio button.

***Step 4.2.1*:** Configure the database search parameters under the "Parameters" tab. Users can also choose to expand their search by utilizing PEAKS PTM and SPIDER. To do this check the "Find unspecified PTMs and common mutations with PEAKS PTM" and "Find more mutations with SPIDER" buttons at the bottom of the window.

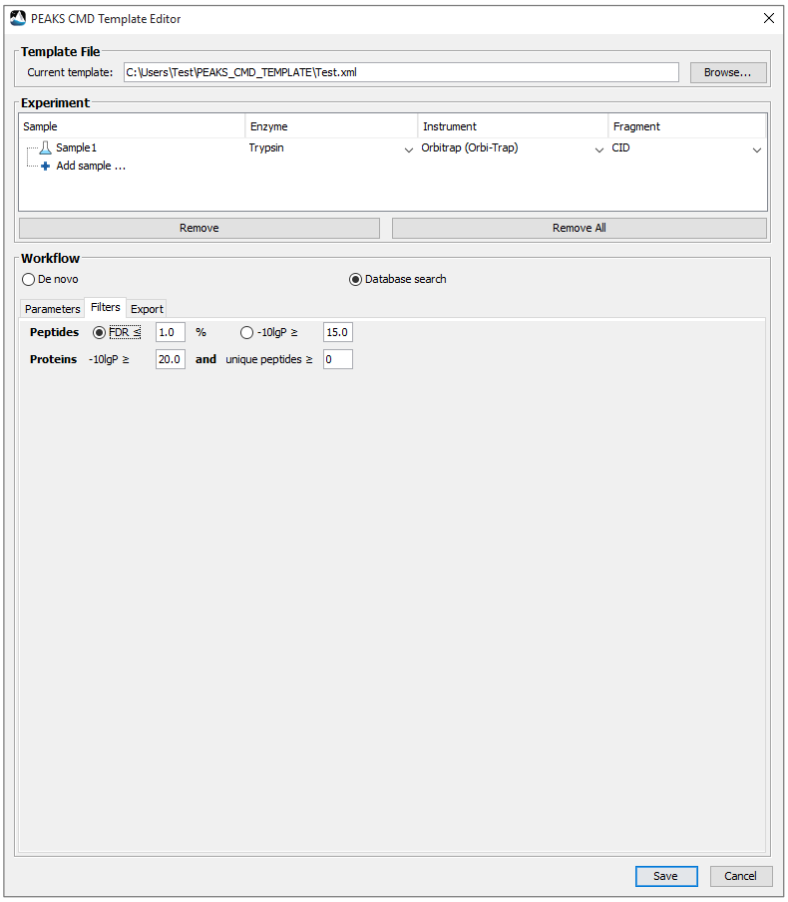


Configure the protein sequence database in the “Database” section under the “Parameters” tab. A protein sequence database is necessary for performing the PEAKS database search. Users can either select one of the existing protein databases created in a previous search or create a new database entry from a FASTA file.

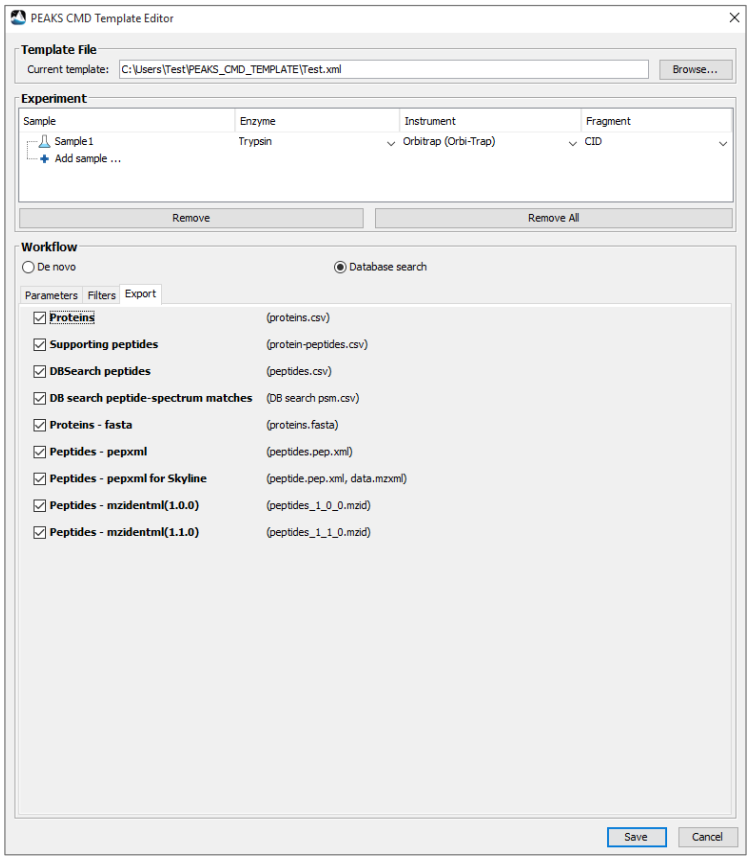
If there is no database entry in the list, users must first click the “View” button and create a database through the prompts in a pop-up “Database Info” dialog by then clicking the “New” button. Otherwise, users can simply select an existing database entry for the subsequent data analysis. Note that users can always click the “View” button to modify an existing database entry in PEAKS CMD.



***Step 4.2.1*:** Set the database search result filter under the “Filters” tab.



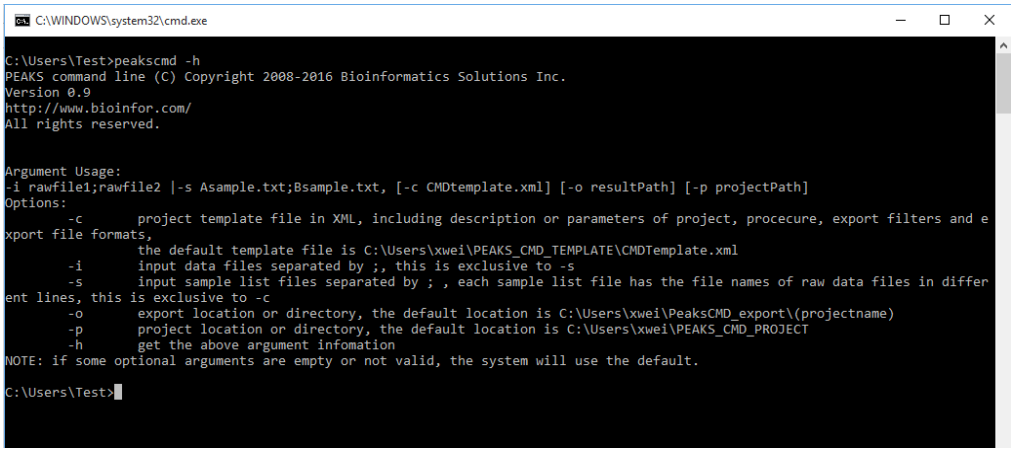
***Step 4.2.2*:** Select the files required for export under the "Export" tab.



***Step 5*:** Click the "Save" button to save the template file and close the TemplateEditor.

5.2 Running PEAKS CMD with a Template File

PEAKS CMD supports six arguments. Running PEAKS CMD with the argument "-h" or "-H" generates the list of six valid arguments as shown below:

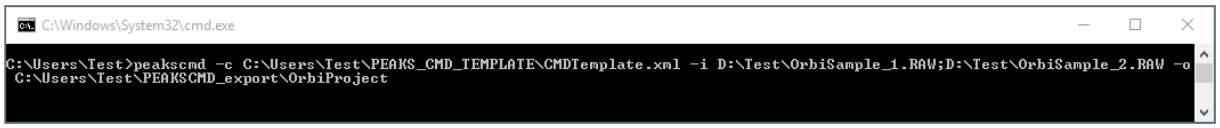


Among the six arguments, -i or -s are mandatory for adding the sample spectra data and are mutually exclusive. A user will then choose one of them and specify the required files or file lists. All other arguments are optional. If the input values for some of the optional arguments are not valid, PEAKS CMD will use the pre-specified default values.

Two examples of using the -i arguments are shown below:

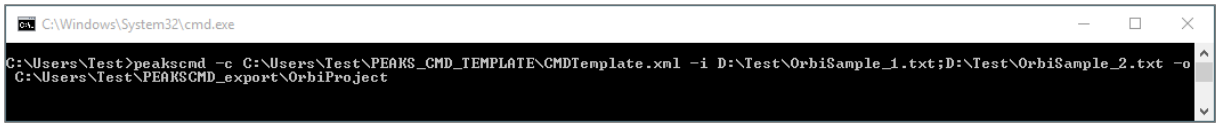
1) Using the -i argument for loading only one data file:

2) Using the -i argument for loading multiple data files:



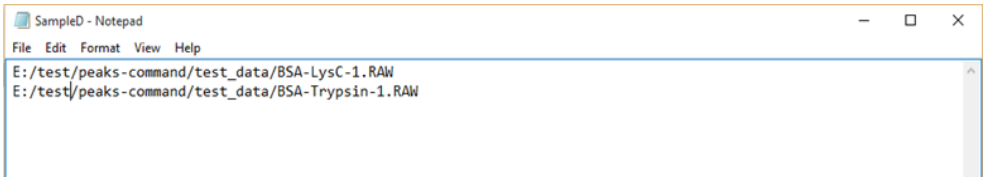
The -s argument requires one or multiple text files in which each line lists an absolute path of a data file. When using the -s argument, the number and the names of the sample files must be identical to the number and names of the samples in template file generation using the TemplateEditor.

An example of using the -s argument with multiple sample files is shown below:



An example of a sample file is shown below:

NOTE: The spectral data files in each sample file are listed line by line, or separated by ";" in a single line.



If values are not specified for the optional arguments the default template file is used automatically, the project is saved into a default folder, and the results is exported into a default folder. The absolute path of the default template file is C:\Users\user_name\PEAKS_CMD_TEMPLATE\CMDTemplate.xml, the default project location for is C:\Users\user_name\PEAKS_CMD_PROJECT, and the default directory of the exporting files is C:\Users\user_name\PeaksCMD_export/project_name.

PEAKS CMD supports the direct loading of RAW files from ThermoFisher, Waters, and Agilent instruments. Multiple types of text-based spectral data files, *eg.* PKL, MGF, mzXML, and mzML, are also supported.