PEAKS® CMD

PROTEOMIC COMMAND LINE SOLUTION

Linux User Guide
December, 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 Linux 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 Linux Operating Systems (Ubuntu, Redhat, Fedora and others). 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 software installation package files to the intended computational resource (pipeline core or computer, or server).

There are three types of PEAKS CMD setup files, DEB, RPM, and a compressed package, for Linux users. The DEB setup file, with .deb as the extension name, is for Ubuntu or Linux distributions with Ubuntu-core (eg. Linux Mint) and it can be installed using "dpkg" command. The RPM setup file, with .rpm as the extension name, is for Red Hat or its other branches or distributions (eg. Fedora and CentOS) and it can be installed using "rpm" command. For some Linux distributions that support both DEB and RPM installation, you can select either of the packages to install PEAKS CMD. There is a third file type (.tar.gz) which is a compressed archive file for the manual installation.

In the event the installation file cannot be found, please contact sales@bioinfor.com. They can also assist with registration key issues.

Step 1: Open a Linux terminal, locate the PEAKS CMD software package directory, and run the following command to install the PEAKS CMD software.

Case 1: Ubuntu or Ubuntu-like Linux users

```
$ sudo dpkg -i peakscmd-{version}-setup.deb
```

Case 2: Red Hat or Red Hat-like Linux users

```
$ sudo dpkg -i peakscmd-{version}-setup.deb
```

Case 3: Any Linux users installing PEAKS CMD using the .tar.gz file

```
$ sudo cp peakscmd-{version}-setup.tar.gz /opt
$ cd /opt
$ sudo tar zxvf peakscmd-{version}-setup.tgz
$ cd PEAKSCMD
$ sudo ./install.sh
```

It is recommended to install the PEAKS CMD into /opt. However, PEAKS CMD can be installed into any other folder for convenience.

NOTE:
1. Users must use the **root** user or **sudo** to install the software and input the exact file name after -i or -ivh, for example, sudo dpkg -i peakscmd-{version}-setup.deb.

2. If there is a version of the PEAKS CMD software on this computer, it may be preferred to uninstall the software first before the new installation using the following commands:

Case 1: Ubuntu or Ubuntu-like users

$sudo dpkg -r peakscmd

Case 2: Red Hat or Red Hat-like users

$sudo rpm -e PEAKSCMD-{version}-1.i386

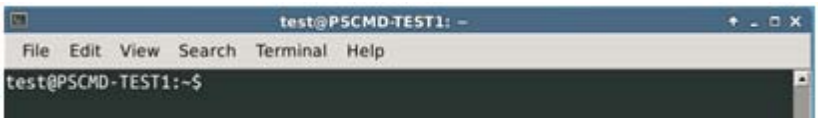
PEAKSCMD-{version}-1.i386 are the package names installed on the system. To identify the version number, run the command "sudo rpm -q –a | grep PEAKSCMD ".

Case 3: Any Linux users installing PEAKS CMD using the .tar.gz file

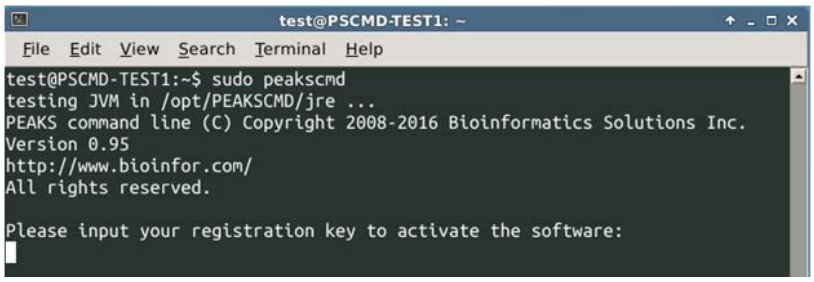
$ sudo rm -r /opt/PEAKSCMD
$ sudo rm -f /usr/bin/peakscmd
$ sudo rm -f /usr/bin/MassCalculator
$ sudo rm -f /usr/bin/PerformanceConfiguration
$ sudo rm -f /usr/bin/TemplateEditor
$ sudo rm -f /usr/bin/ProjectConverter

4. Registering PEAKS CMD

A Linux terminal must be launched to run PEAKS CMD. Typically, a Linux terminal will start to run as shown right:

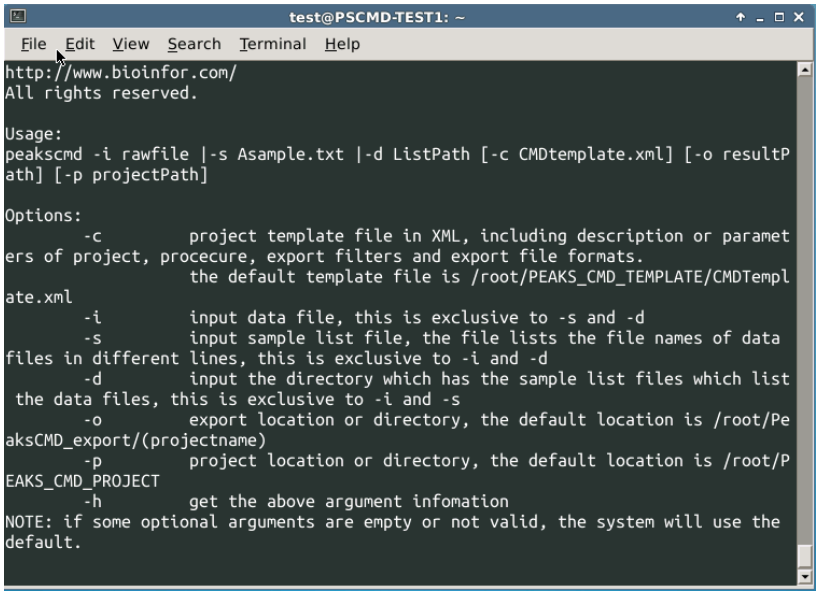


Type "peakscmd" in the terminal to run PEAKS CMD. The first time PEAKS CMD is run, it will ask for a registration key to activate the software. Run the command "peakscmd" by the root or a sudo user.



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

Usage information will be shown if the registration is successful.

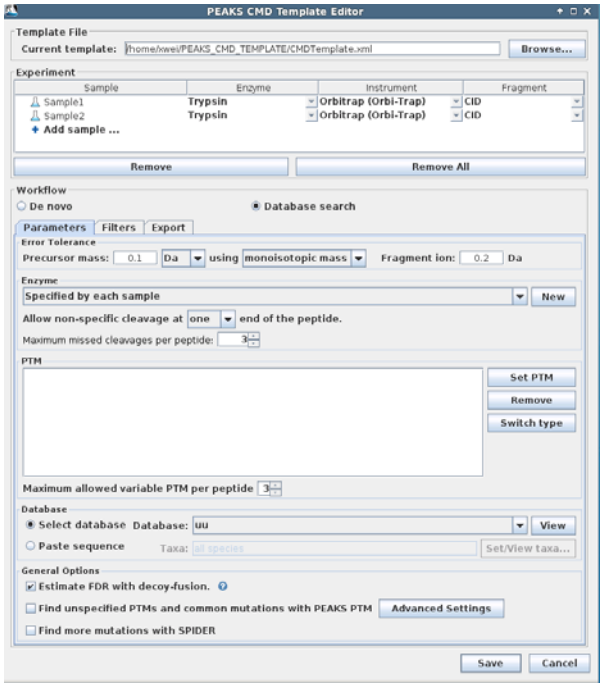


5. Using PEAKS CMD

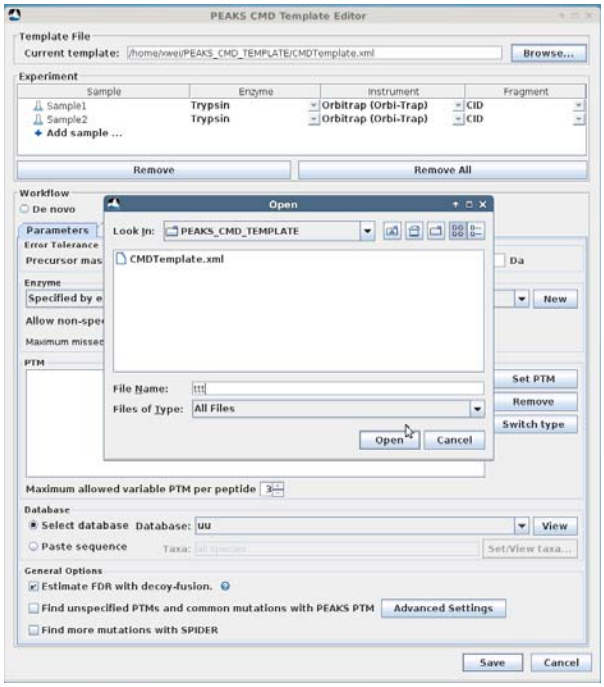
Performing protein/peptide identification using PEAKS CMD requires two steps: 1) generating a template file and 2) running PEAKS CMD software with a template file.

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 the command "TemplateEditor" on terminal. The editor will appear as presented:



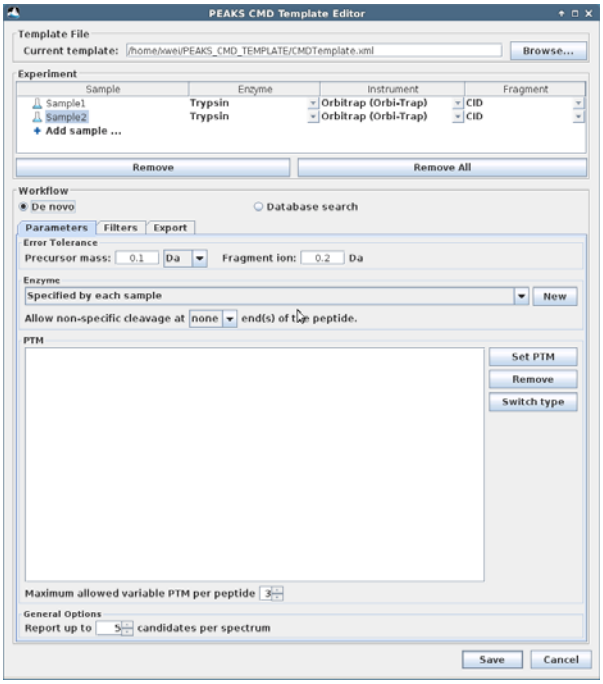
Step 1: Create or edit the template file in the "Template File" section. The path shown in the "Current Template" text field is the default path to save a template file. Users can also save template files in a preferred directory and with different files name by clicking "Browse…".



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.

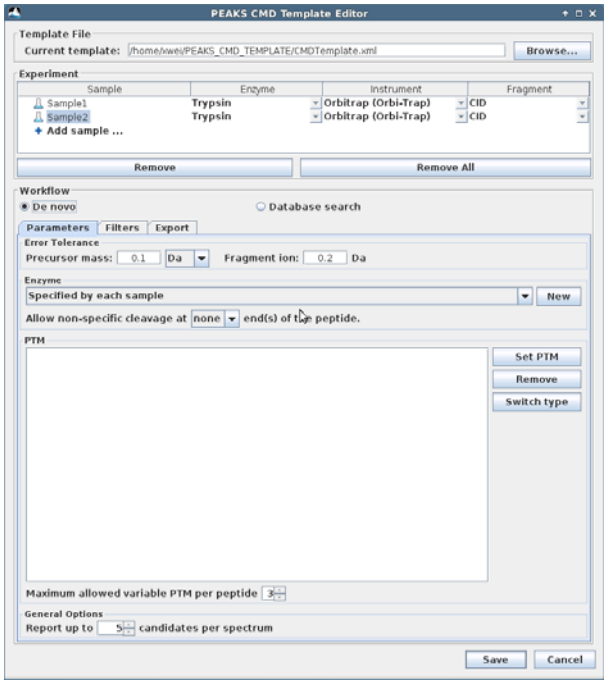
NOTE: Adding sample files in this step is not necessary. Users only need to set the sample numbers and names for the experiment.

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.

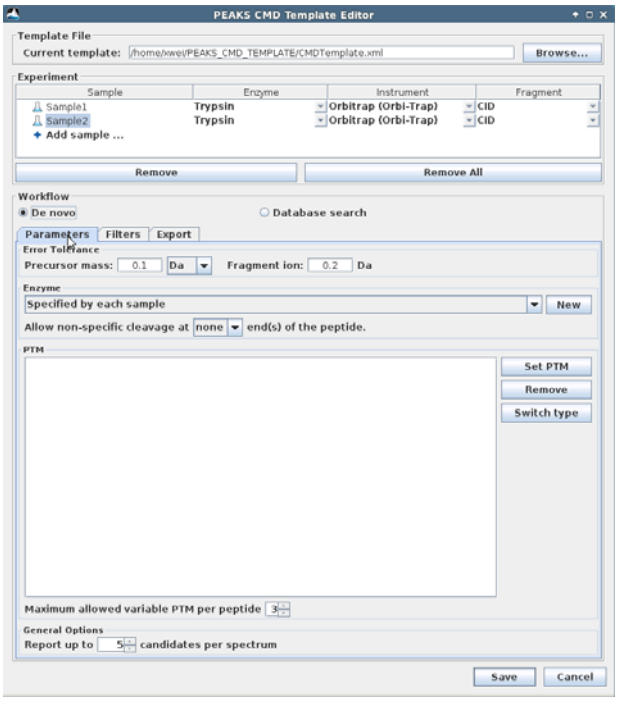


Step 4: Configure PEAKS CMD task parameters. Typically, *de novo* sequencing and database searching require slightly different parameter configurations. The following steps demonstrate how to configure the parameters required in *de novo* sequencing and database searching, respectively.

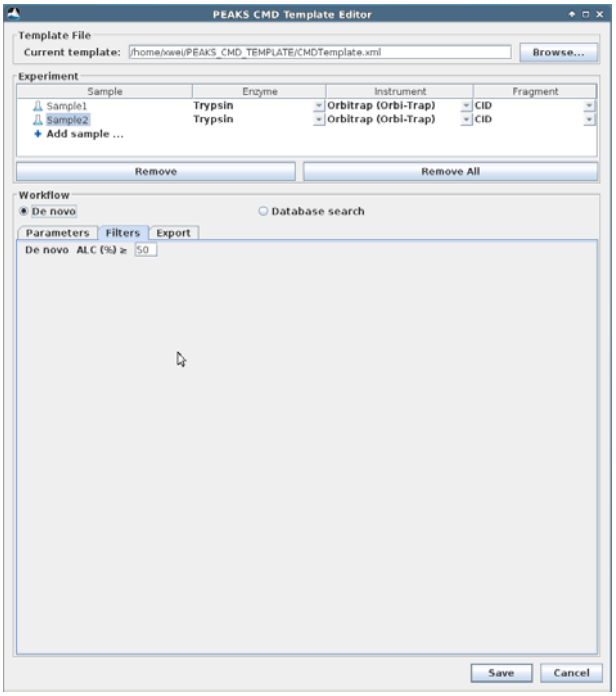
Option 4.1: Check the "De novo" radio button.



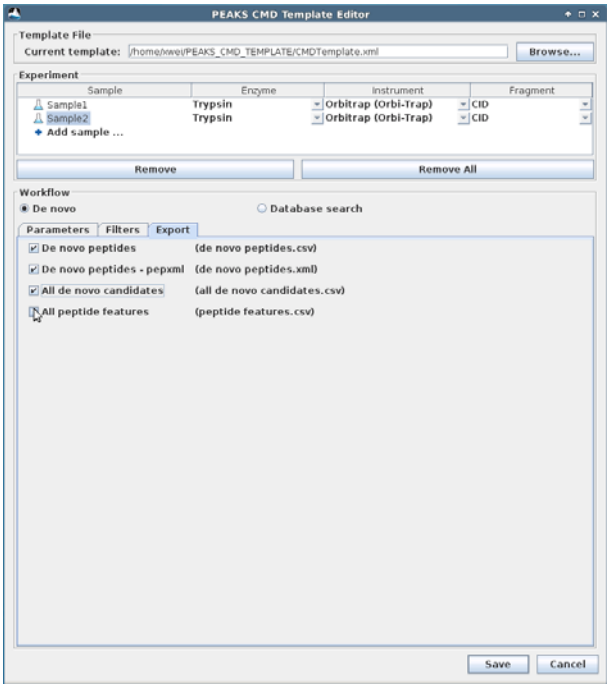
***Step 4.1.1*:** Configure the *de novo* sequencing parameters under the "Parameters" tab.



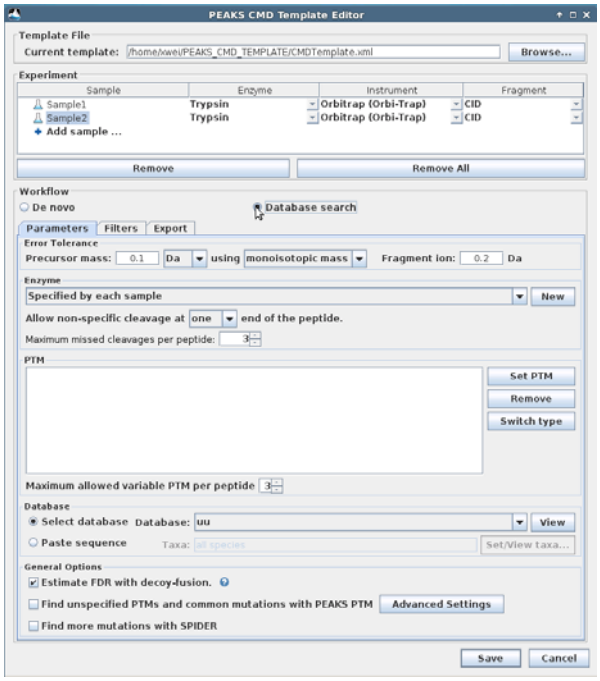
***Step 4.1.2*:** Set the *de novo* sequencing result filter under the "Filters" tab.



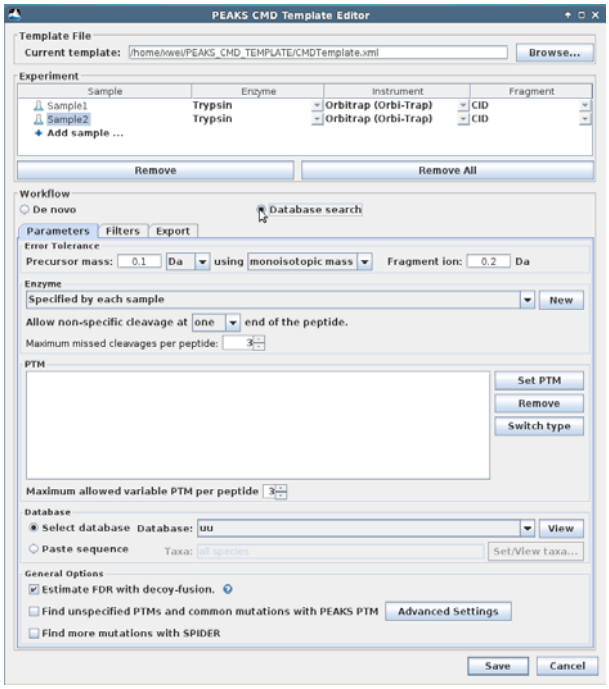
Step 4.1.3. Select the files required for export under the "Export" tab.



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

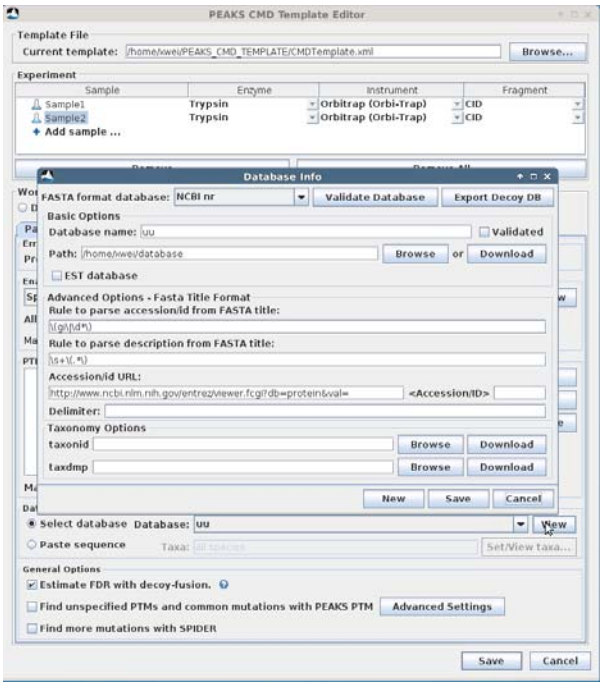


***Step 4.2.1*:** Configure the database search parameters under the "Parameters" tab. Users can also choose to expand their search by utilizing PEAKS PTM and SPIDER. To do this check the "Find unspecified PTMs and common mutations with PEAKS PTM" and "Find more mutations with SPIDER" buttons at the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

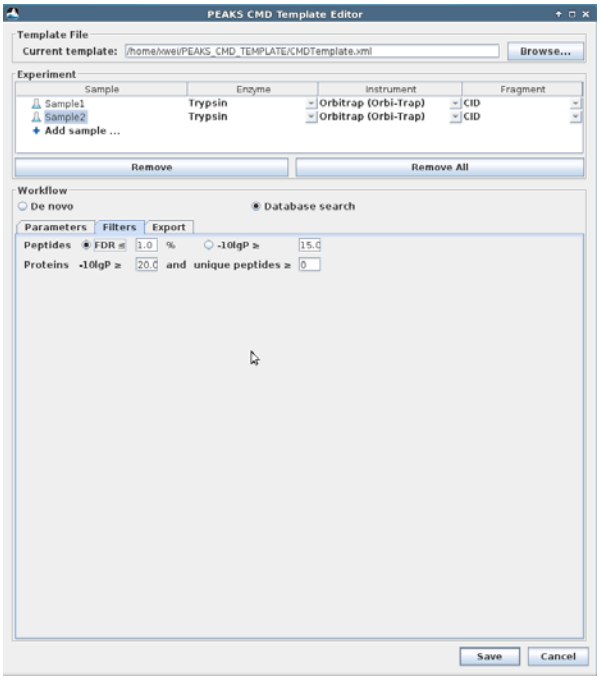


Configure the protein sequence database in the "Database" section under the "Parameters" tab. A protein sequence database is necessary for performing the PEAKS database search. Users can either select one of the existing protein databases created in a previous search or create a new database entry from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

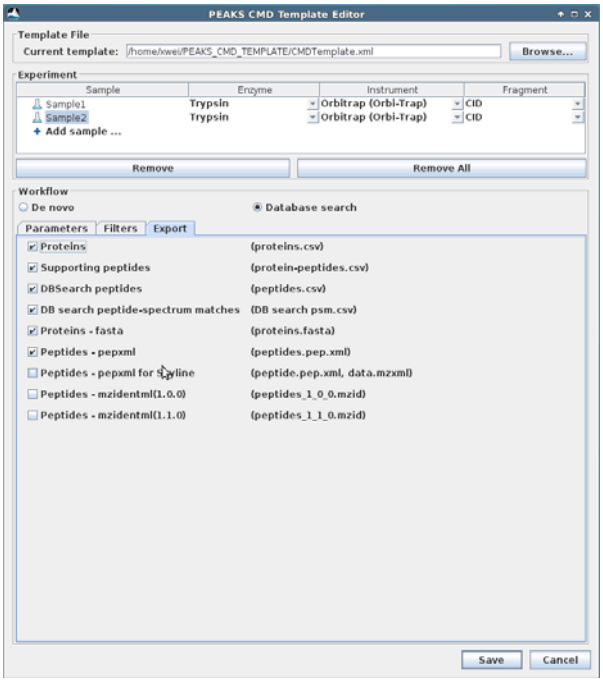
If there is no database entry in the list, users must first click the "View" button and create a database through the prompts in a pop-up "Database Info" dialog by then clicking the "New" button. Otherwise, users can simply select an existing database entry for the subsequent data analysis. Note that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

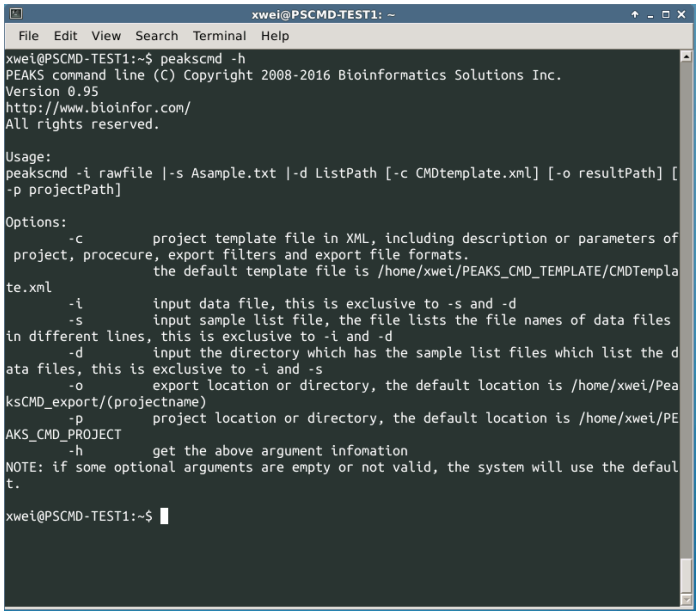


***Step 5*:** Click the "Save" button to save the template file and close the TemplateEditor.

NOTE: If the Linux server or computer does not have a GUI or desktop environment, it is recommended to run the "TemplateEditor" on a computer (Linux or Windows) with a GUI to generate the template file, then copy or upload the template file into the Linux server and use it. In step 4.2.1, the absolute path must be input for database entry on Linux server. This is also true on the Windows version TemplateEditor. Be sure to use the separator "/" for Linux, not "\" which is used for Microsoft Windows.

5.2 Running PEAKS CMD with a Template File

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



Among the seven arguments, -i or -s or -d 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. The directory followed by -d must contain the sample list files with extensions .txt, which list the data files in different lines.

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 and -s arguments are shown below:

- Using the -i argument for loading only one data file:
  $ peakscmd -c /home/test/template/exp1.xml -i /home/test/data/protein1.mzML

- Using the -s argument for loading multiple data files for one sample:
  $ peakscmd -c /home/test/template/exp1.xml -s /home/test/slist/exp1.txt

  Where exp1.txt is a text-based file containing the absolute paths of mass spectral files line-by-line as shown below:
  /home/test/data/pr1.mzML
  /home/test/data/pr2.mzML
  /home/test/data/pr4.mzML

When using the -d argument, the number of the sample files in the directory following -d must be identical to the number of the samples in template file generated by the TemplateEditor.
An example of using the -d argument with multiple sample files is shown below:

$ peakscmd -c /home/test/template/exp_multiple.xml -d /home/test/slist

Listing all the files in the folder /home/test/slist shows the following:
$ ls -l /home/test/slist
-rwxrw-rw- 1 test test 7088 Dec 4 13:48 Sample1.txt
-rwxrw-rw- 1 test test 5034 Dec 4 13:48 Sample2.txt

Each sample file, *eg.* Sample1.txt, may contain multiple mass spectral files:
/home/test/data/T1_pr12.mzML
/home/test/data/T2_pr25.mzML
/home/test/data/T1_pr33.mgf

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 $HOME/PEAKS_CMD_TEMPLATE/CMDTemplate.xml, the default project location for is $HOME/PEAKS_CMD_PROJECT, and the default directory of the exporting files is $HOME/PeaksCMD_export/project_name.
PEAKS CMD supports multiple types of text-based spectral data files, *eg.* PKL, MGF, mzXML, and mzML.