PDV user's manual

Version 1.5.2

Table of Contents

1 Introduction

PDV (https://github.com/wenbostar/PDV) is a lightweight Java-based visualization tool that enables intuitive and fast exploration of diverse, large-scale proteomics datasets on standard desktop computers in both GUI and command line modes. If you have any questions, suggestions or remarks, please join us on Gitter (https://gitter.im/PDV-public/Lobby). For specific bug reports or issues please use the Github issues tracker (https://github.com/wenbostar/PDV/issues).

2 Required resources

Software download: please download the latest version of PDV from the following website: https://github.com/wenbostar/PDV/releases.

Java version: 1.8 or later

OS: Windows, Linux and Mac OS

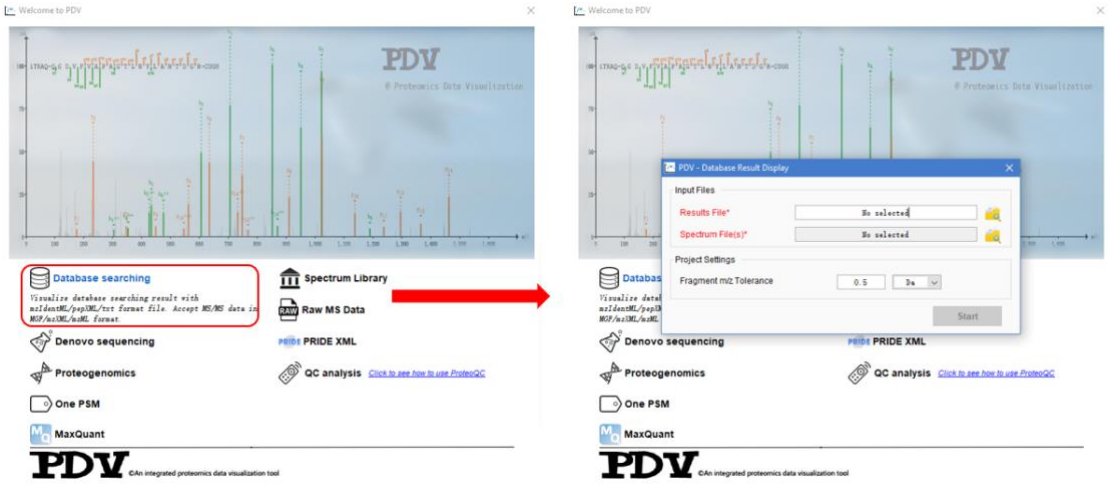
Hardware: 2 CPUs (more is better), 4 Gb memory

Input data for testing: Please download the testing data from the following website: http://pdv.zhang-lab.org/data/download/pdv_upload.tar.gz.

3 Using PDV in GUI mode

3.1 Database searching result visualization

Users can open the panel for database searching result visualization as shown below:



Two types of files are required as input:

(1) A peptide identification result file. Supported identification result format:

- mzIdentML: http://www.psidev.info/mzidentml
- pepXML: http://tools.proteomecenter.org/wiki/index.php?title=Formats:pepXML
- txt: a tab-delimited file like the table shown below:

peptide	modification	spectrum_title	charge	pep_mass	mz
VAPQNDSFGTQLPPMHQQQR	-	113658	4	2278.0913	570.535522
GKGAAAAAAASGAAGGGGG GAGAGAPGWGR	-	43571	3	2266.09514	756.365112
VGAACPAPGTGSGPLR	Carbamidomethyl of C@5[57.0215]	110145	3	1466.73	489.911041

Please note if the input identification file is a txt file, then the input MS/MS file must be an MGF format file. The “spectrum_title” is referred to the title in the MGF file. PDV requires a specific format of modification as shown in the above table. The name of the modification can be the modification name existing in unimod (http://www.unimod.org/modifications_list.php) database or any use-defined modification name. If there is no modification for a peptide, the value of the modification column should be assigned as “-”. If there are multiple modifications in the same peptide, modifications must be separated by “;”, such as “Carbamidomethyl of C@5[57.0215];Oxidation of M@2[15.994915]”. For N-term modification the position should be 0, such as “Acetyl of N-term@0[42.010565]”. As For C-term modification the position should be (length of peptide + 1), such as “Homoserine of C-term@11[-29.992806]” for peptide VGAACPAPGR.

Supported search engines (protein identification software):

MS-GF+, MyriMatch, X!Tandem, OMSSA (convert OMSSA raw result to mzIdentML format using MzidLib: https://github.com/PGB-LIV/mzidlib), Crux/Tide, Comet, Mascot, IPeak and

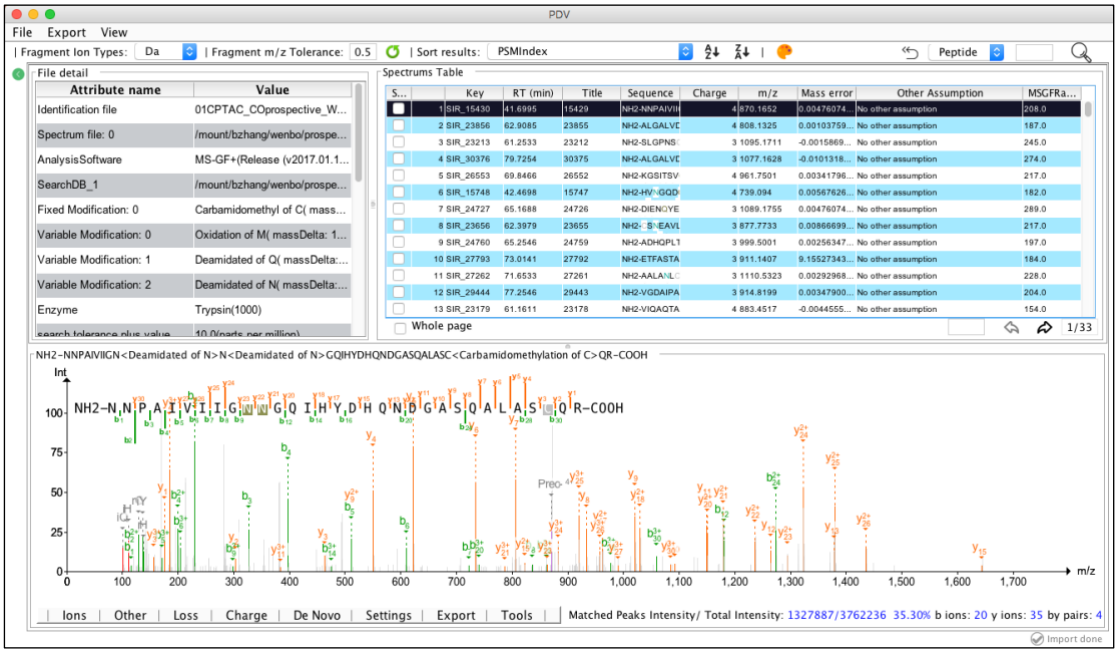
MSFragger. For each software, the input formats supported are listed at https://github.com/wenbostar/PDV#database-searching. For each supported input format, example input files and the version of corresponding tool used to generate the identification result files are also available there.

(2) An MS/MS data file. Supported MS/MS data format:

- MGF: http://www.matrixscience.com/help/data_file_help.html#GEN
- mzML: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013463/
- mzXML: http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML

Please note that the input MS/MS data file must be the same with that is used to generate the input identification result file.

After the two files are loaded, a panel like below will be shown:



Below please find detailed description of the above result panel:

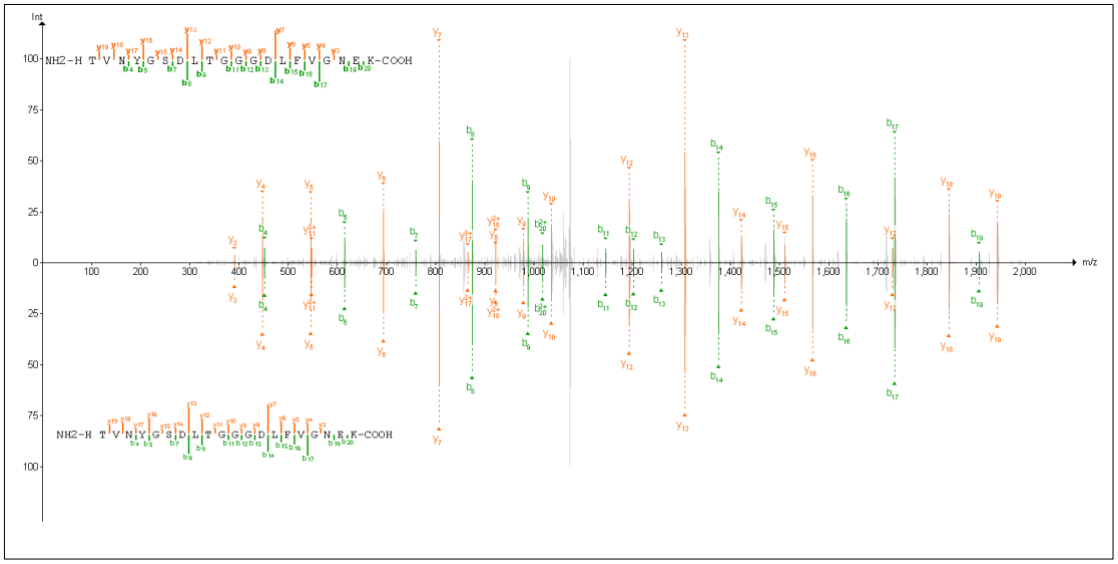
① Sort the whole PSM table (across all pages) according the column selected by users; the sorting can be in increasing model or decreasing model by clicking or .

② Set the mass tolerance window for MS/MS fragment ion matching;

③ Peptide identification parameters from the input identification file;

④ Spectrum annotation figure panel;

⑤ Set detailed parameters for spectrum peptide matching. Users can export annotated spectrum using the “Export” function. In addition, users can use the function in “Tools” menu to import a new spectrum to generate figure as shown below:



⑥ Set modification amino acid color;

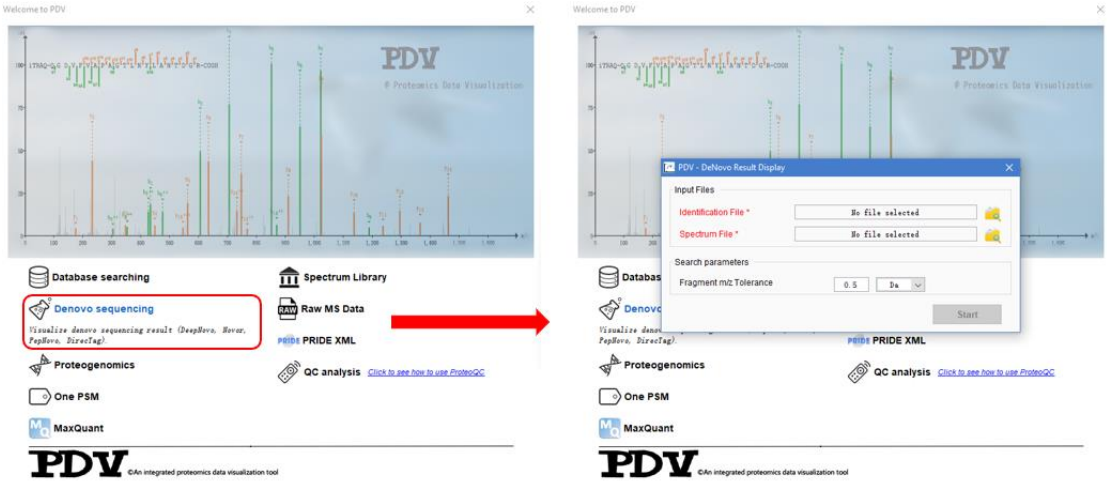
⑦ Filter rows in whole PSM table (across all pages) by peptide sequence or spectrum ID;

⑧ PSM table containing all the peptide spectrum match data;

⑨ Peak annotation information.

3.2 Denovo sequencing result visualization

Users can open the panel for De novo searching result visualization as shown below:



Two types of files are required as input:

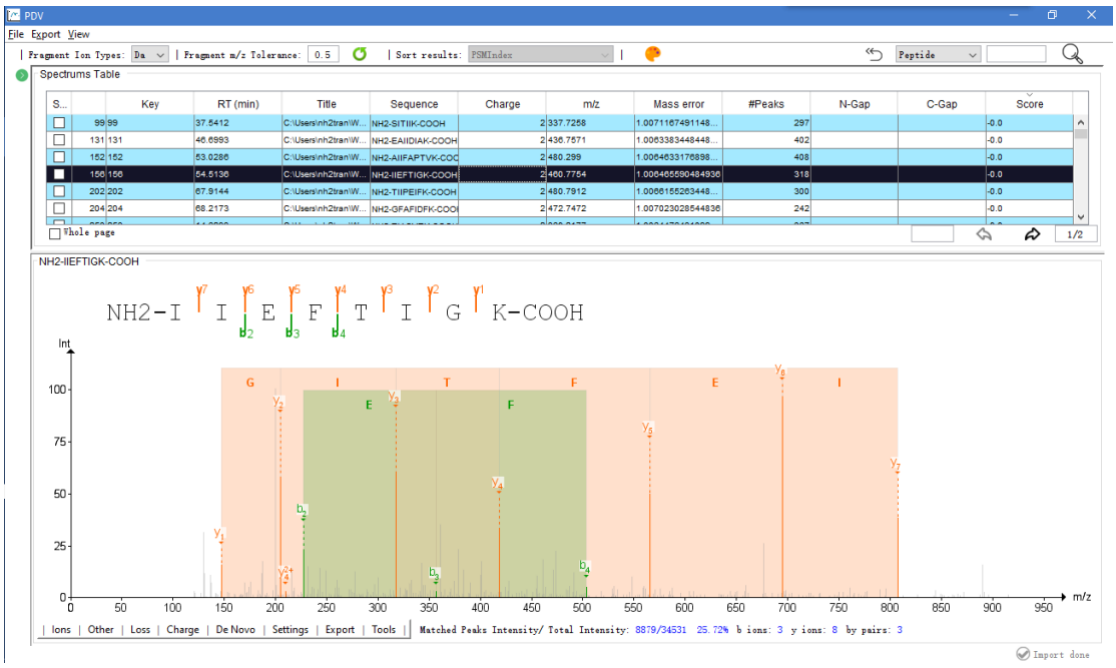
(1) A peptide identification result file. Supported De novo software results:

- DeepNovo: https://github.com/nh2tran/DeepNovo
- Novor: https://www.rapidnovor.com/download/
- PepNovo: https://github.com/jmchilton/pepnovo
- Directag: http://proteowizard.sourceforge.net/downloads.shtml, select Bumbershoot*.

(2) An MS/MS data file. Supported MS/MS data format:

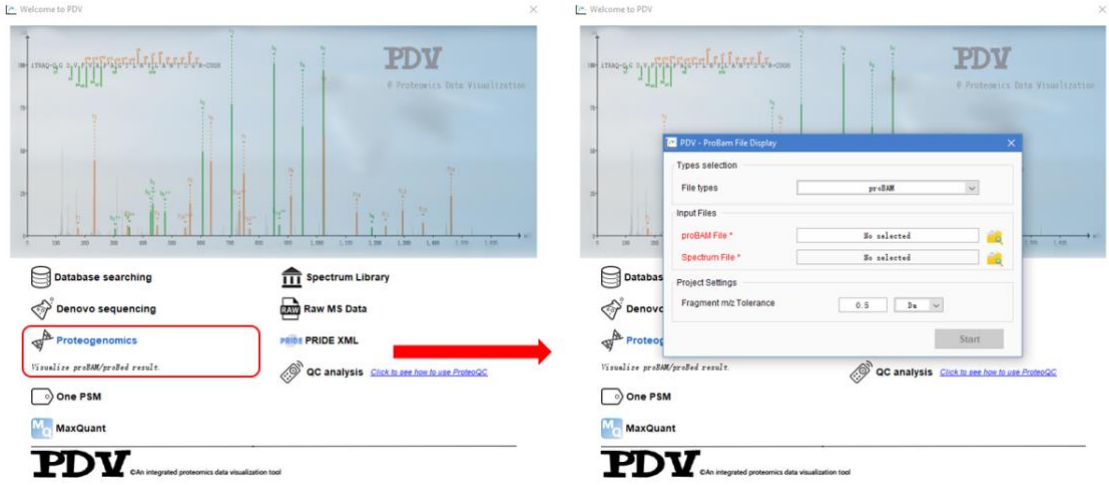
- MGF: http://www.matrixscience.com/help/data_file_help.html#GEN

After the two files are loaded, a panel like below will be shown:



3.3 Proteogenomics data visualization

Users can open the panel for proteogenomics file visualization as shown below:



Two types of files are required as input:

(1) A proteogenomics data file which includes peptide spectrum match information. Supported file format:

- proBAM: http://www.psidev.info/proBAM

- proBed: http://www.psidev.info/proBed. If the input is a proBed file, the mzIdentML file which is used to generate the proBed is also required.

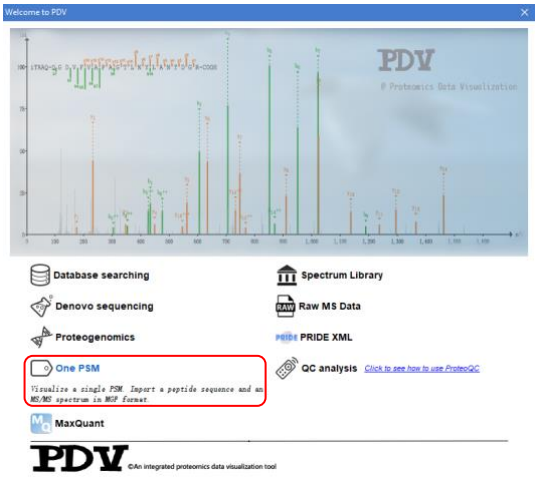
(2) An MS/MS data file. Supported MS/MS data format:

- MGF: http://www.matrixscience.com/help/data_file_help.html#GEN
- mzML: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013463/
- mzXML: http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML

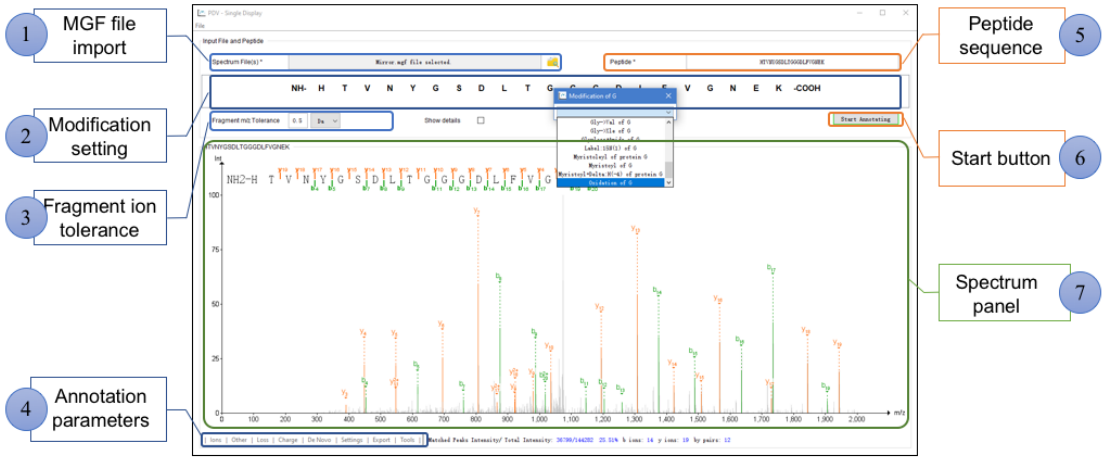
After the files are loaded, a panel like database searching results visualization will be shown.

3.4 One PSM visualization

Users can open the panel for single PSM visualization as shown below:



When users click “One PSM” as shown in the above panel, a panel like below will be shown.



Two kinds of information are required as input:

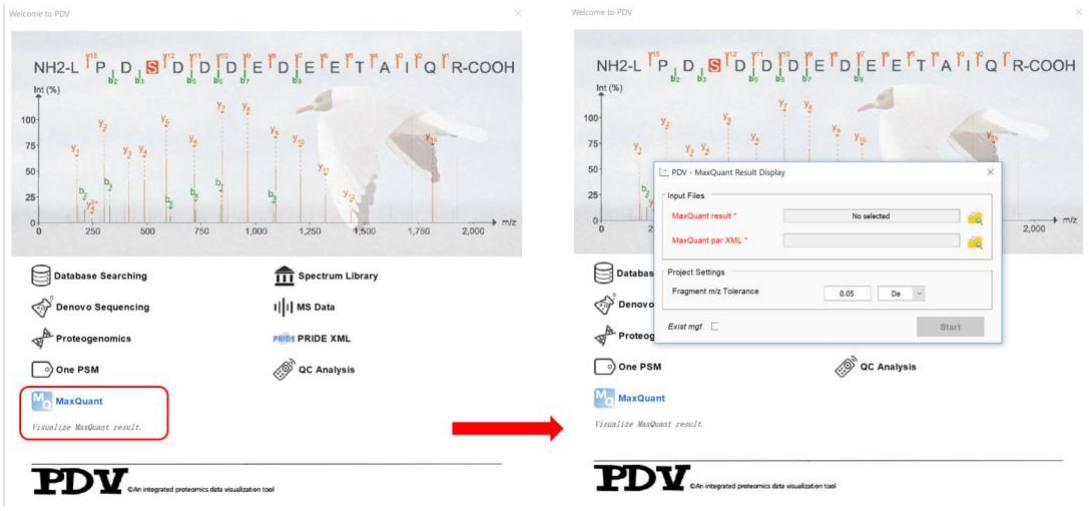
(1) An MGF file containing only one spectrum;

(2) Peptide sequence.

Modifications can be set by clicking the corresponding amino acid. Currently only support one modification for one modification.

3.5 MaxQuant result visualization

Users can open the panel for database searching result visualization as shown below:

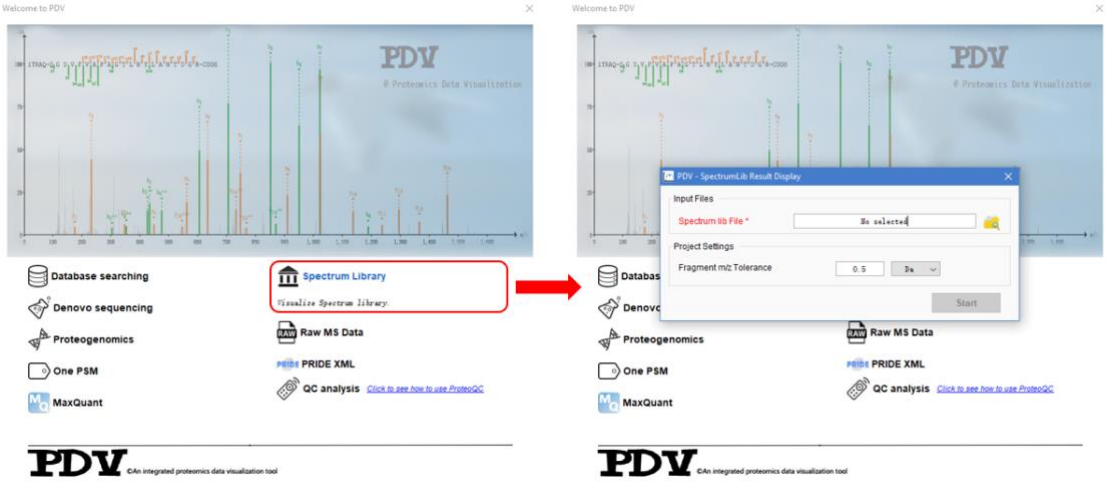


Two inputs are required: (1) one folder named **combined** which is generated by MaxQuant and (2) **mqpar.xml** file generated by MaxQuant.

If this is the first time you import the **combined** folder, PDV will automatically read MS/MS data in the **combined** folder and generate mgf files in the **combined** folder. If it's not the first time you import this **combined** folder, then you can select "Exist mgf" to skip the step of generating mgf files.

3.6 Spectrum library visualization

Users can open the panel for spectrum library visualization as shown below:



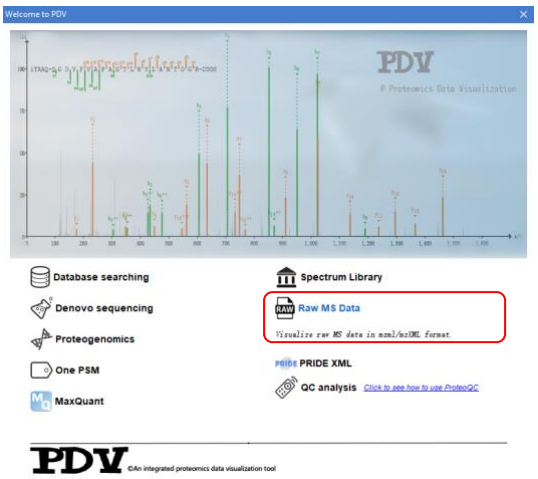
One folder containing spectrum library data is required as input:

Support format is splib: http://www.peptideatlas.org/speclib/.

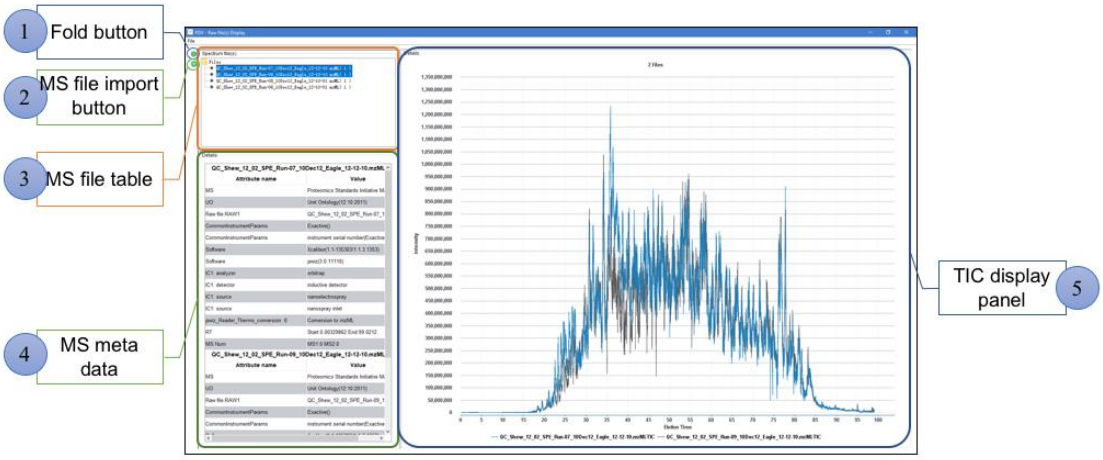
After the folder is loaded, a panel like database searching results visualization will be shown.

3.7 Raw MS data visualization

Users can open the panel for raw files visualization as shown below:



When MS/MS data is loaded, a panel like below will be shown:



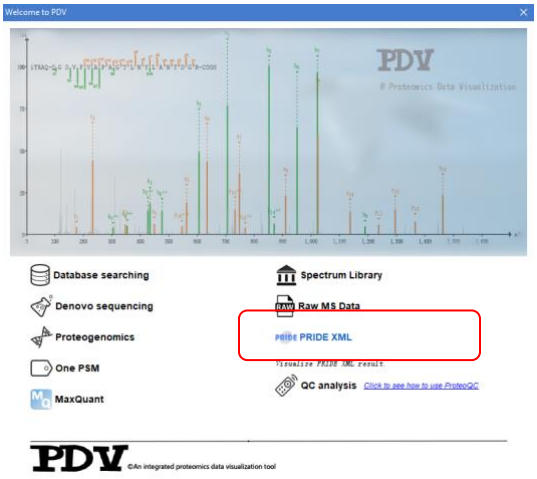
Two kinds of raw files format are accepted:

- mzML: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3013463/
- mzXML: http://tools.proteomecenter.org/wiki/index.php?title=Formats:mzXML

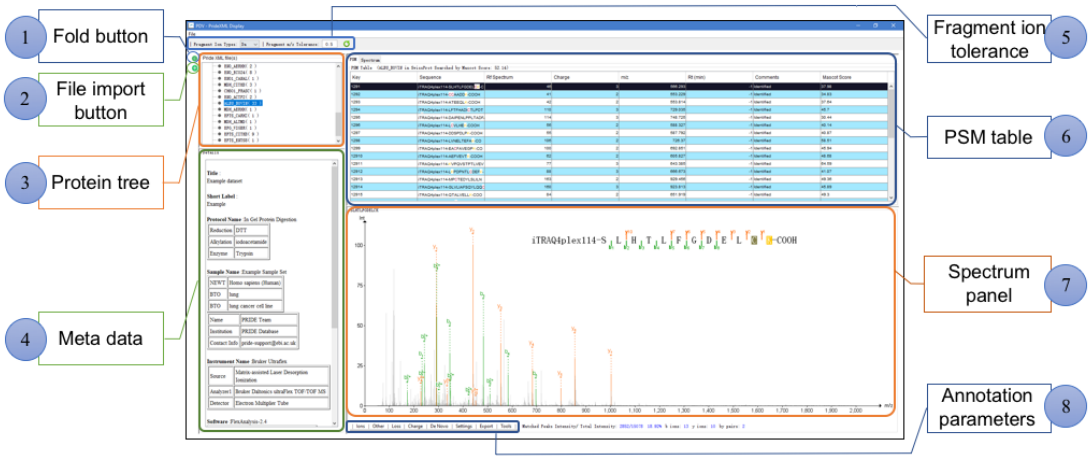
Users can import multiple MS/MS file.

3.8 PRIDE XML visualization

Users can open the panel for PRIDE XML files visualization as shown below:

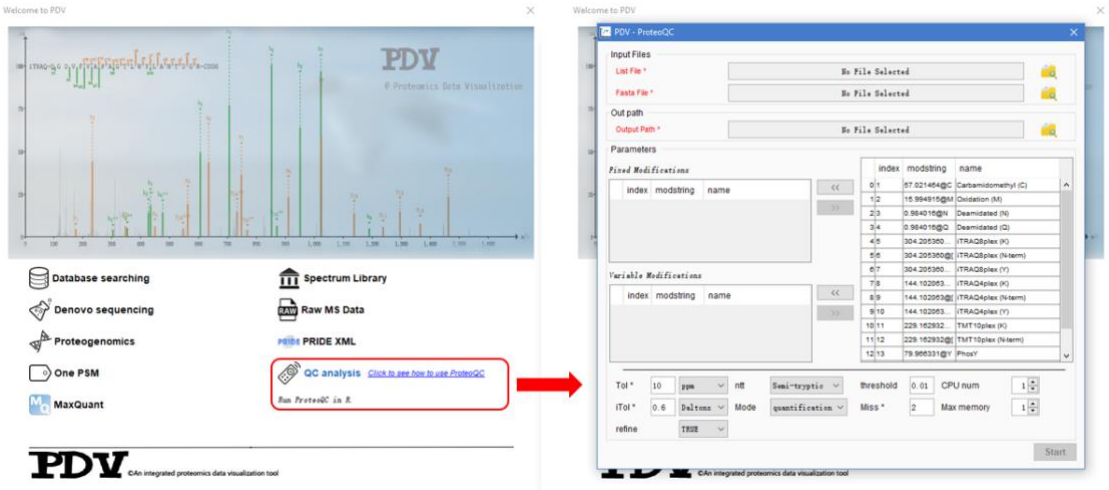


When data is loaded, a panel like below will be shown:



3.9 Proteomics data QC analysis

Users can open the panel for data quality analysis by using proteoQC (http://bioconductor.org/packages/proteoQC/) as shown below. If users want to use this function, R and proteoQC must be installed firstly. PDV will automatically generate an R script based on the inputs and then call R to run the R script. Finally, a QC report will be generated by proteoQC.



4 Using PDV in command line mode

Using following command line will print the command line parameters:

java -jar PDV-1.0.5.jar -h

All the command line parameters are shown as below:

```
 $ java -jar PDV-1.0.5.jar -h
usage: Options
 -a <arg>    Error window for MS/MS fragment ion mass values. Unit is Da.
             The default value is 0.5.
 -ah         Whether or not to consider neutral loss of H2O.
 -an         Whether or not to consider neutral loss of NH3.
 -c <arg>    The intensity percentile to consider for annotation. Default
             is 3 (3%), it means that the peaks with intensities >= (3% *
             max intensity) will be annotated.
 -fh <arg>   Figure height. Default is 400
 -ft <arg>   Figure type. Can be png, pdf or tiff.
 -fu <arg>   The units in which 'height'(fh) and 'width'(fw) are given.
             Can be cm, mm or px. Default is px
 -fw <arg>   Figure width. Default is 800
 -h          Help
 -help       Help
 -i <arg>    A file containing peptide sequences or spectrum IDs. PDV will
             generate figures for these peptides or spectra.
 -k <arg>    The input data type for parameter -i (Spectrum ID: s, peptide
             sequence: p).
 -o <arg>    Output directory.
 -pw <arg>   Peak width. Default is 1
 -r <arg>    Identification file.
 -rt <arg>   Identification file format (mzIdentML: 1, pepXML: 2, proBAM:
             3, txt: 4, maxQuant: 5).
 -s <arg>    MS/MS data file
 -st <arg>   MS/MS data format (mgf: 1, mzML: 2, mzXML: 3).
```

Please find some example command lines in the following table:

Identification file	MS/MS format	Spectrum key format	Command line examples

mzIdentML	MGF	Spectrum title	Java -jar PDV-1.0.5.jar -r E:\Example.mzid -rt 1 -s E:\Example.mgf -st 1 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
	mzML	Scan number	Java -jar PDV-1.0.5.jar -r E:\Example.mzid -rt 1 -s E:\Example.mzML -st 2 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
	mzXML	Scan number	Java -jar PDV-1.0.5.jar -r E:\Example.mzid -rt 1 -s E:\Example.mzXML -st 3 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
proBAM	MGF	Spectrum title	Java -jar PDV-1.0.5.jar -r E:\Example.bam -rt 3 -s E:\Example.mgf -st 1 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
	mzML	Scan number	Java -jar PDV-1.0.5.jar -r E:\Example.bam -rt 3 -s E:\Example.mzML -st 2 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
	mzXML	Scan number	Java -jar PDV-1.0.5.jar -r E:\Example.bam -rt 3 -s E:\Example.mzXML -st 3 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
MaxQuant	-	Scan number	Java -jar PDV-1.0.5.jar -r E:\combined -rt 5 -s . -st 1 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
pepXML	mzML	Scan number	Java -jar PDV-1.0.5.jar -r E:\Example.pepXML -rt 2 -s E:\Example.mzML -st 2 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf
text	mgf	Spectrum title	Java -jar PDV-1.0.5.jar -r E:\Example.txt -rt 4 -s E:\Example.mgf -st 1 -i E:\ExampleIndex.txt -k s -o E:\ExampleFolder -a 0.5 -c 5 -pw 1 -fw 800 -fh 400 -fu px -ft pdf

Please note if the input file for identification is a tab-delimited file, the format of this file must be like below:

spectrum_title	**peptide**	**charge**	**modification**
Spectrum_index_1	LKVNVTSK	2	Pyridylacetyl of K@8[119.0371]
Spectrum_index_2	AVAKTHPDKLPESLSLENK	2	Xlink:DMP of K@19[122.0844]
Spectrum_index_3	LCLDVLKTNWSPALQLR	3	Carbamidomethyl of C@2[57.0215]; Iodoacetanilide of K@7[133.0528]
Spectrum_index_4	FNKNEATEMPFR	2	Carboxyethyl of K@3[72.0211]

Please note if you get an error like below in Linux system, it means that your X11 environment doesn't work for your terminal. Please set up X11 environment before you run PDV in command line mode.

```
java -jar PDV-1.0.0.jar --help
Exception in thread "main" java.awt.HeadlessException:
No X11 DISPLAY variable was set, but this program performed an operation which requires it.
at java.awt.GraphicsEnvironment.checkHeadless(GraphicsEnvironment.java:204)
at java.awt.Window.(Window.java:536)
at java.awt.Frame.(Frame.java:420)
at java.awt.Frame.(Frame.java:385)
at javax.swing.JFrame.(JFrame.java:189)
at PDVCLI.PDVCLIMainClass.(PDVCLIMainClass.java:173)
at PDVGUI.gui.PDVMainClass.main(PDVMainClass.java:247)
```