FragPipe-PDV tutorial

*Kai Li, Daniel Polasky*

07/01/2024

- Introduction
- Run FragPipe-PDV
- Inside FragPipe-PDV
- Troubleshooting
- Update FragPipe-PDV in FragPipe manually.
- How to run FragPipe-PDV in CMD

Introduction:

FragPipe-PDV is a customized version of PDV for FragPipe. I changed the GUI layout, optimized the database structure, and implemented new functions fitting FragPipe results especially.
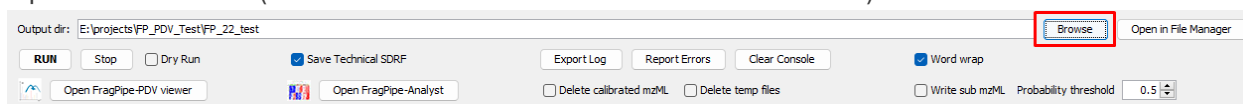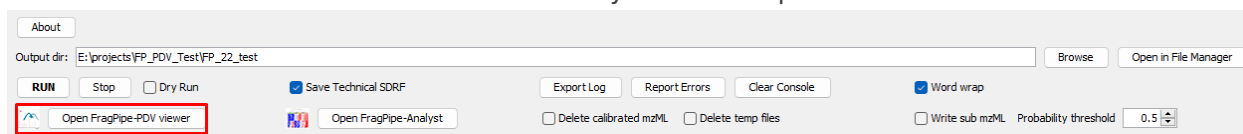
Run FragPipe-PDV:

1. Right after FragPipe search:
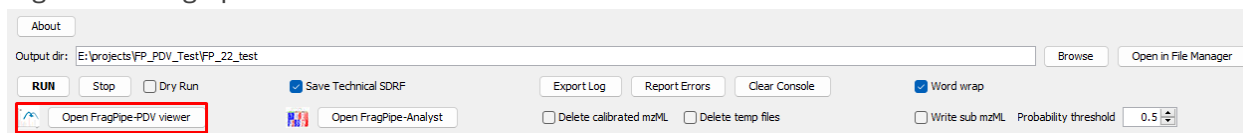
   Click the "Open FragPipe-PDV viewer" button directly.

2. Open an old search (no modification of the result folder and raw files):

   Click "Browse" button to select the result folder you want to open.

   Click the "Open FragPipe-PDV viewer" button directly.

3. Download the FragPipe search result folder online or move from the other places:
   It's a little tricky to open a modified result folder because the FP-PDV viewer uses the "fragpipe-files.fp-manifest" file to find the local paths to the LC-MS files on the machine. If the results folder has been modified, downloaded from an external source, etc., these paths need to be updated to point to all the LC-MS files used in the FragPipe run on the local machine.
   Firstly, please make sure you have at least following files in the result folder plus raw files (.mzML file): "protein.tsv", "protein.fas", "peptide.tsv", "psm.tsv", "fragpipe-files.fp-manifest", and "msbooster_params.txt" if you ran MSBooster.
   Second, open the "fragpipe-files.fp-manifest" file using any text editor. (p.s. you could save the original "fragpipe-files.fp-manifest" file as a copy)
   Third, update the raw files path in the file:

   Fourth, save the "fragpipe-files.fp-manifest".
   Lastly, open FragPipe-PDV following point 2.

Inside FragPipe-PDV:

The main FragPipe-PDV looks like the following:

There are five main panels:

1. "Experiments" panel:

If the user defines multiple experiments in FragPipe "Workflow" panel, the experiments will be shown here in each row. By default, only the first experiment will be selected for visualization. If the user wants to see all experiments results together, please check all experiments in the first column.

2. "Proteins" panel:

This panel shows all proteins identified in the experiment(s) selected in the "Experiments" panel. The contents of this panel are from "protein.tsv" file. By default, only PSMs from the one protein selected will be shown in the "PSM" panel.

How to show PSMs for multiple proteins?
If the user wants to see PSMs from multiple proteins, first check all interested proteins in the first "Selected" column; then click "Show all selected" button.

For example, PMSs from all three selected proteins will be shown in the "PSMs" panel.

How to show PSMs for all proteins?
If the user wants to see PSMs from all identified proteins, first check the "All proteins" box; then click "Show all selected" button.

How to format the table columns?

Since we got all information from the "protein.tsv" file, there are too many columns to show in one page. We design a button on the top right corner letting users to select which columns shown in the panel:

One table will be poped out once clicking the button. The user can check the columns they want to visualize.

How to sort the data?

There are two ways to sort the data in the table:

- Click the column name directly on the table to sort the current page: (**ONLY CURRENT PAGE, NOT WHOLE DATA**)

Selected	Experi...	Protein	Gene	ProteinDescripti...	TotalPeptides	
☐	DDA_Lib	contam_sp|P00330|...	ADH1	Alcohol dehydrogen...	4	
◼	DDA_Lib	contam_sp|P02769|...	ALB	Albumin	4	
☐	DDA_Lib	sp|A0A0B4J2D5|GA...	GATD3B	Putative glutamine a...	3	
☐	DDA_Lib	sp|A0A2R8Y4L2|RA...	HNRNPA1L3	Heterogeneous nucl...	9	
☐	DDA_Lib	sp|A0AVT1|UBA6_H...	UBA6	Ubiquitin-like modifi...	2	
☐	DDA_Lib	sp|A0FGR8|ESYT2_...	ESYT2	Extended synaptota...	4	
☐	DDA_Lib	sp|A0JNW5|BLT3B_...	BLTP3B	Bridge-like lipid tran...	4	
☐	DDA_Lib	sp|A1X283|SPD2B_...	SH3PXD2B	SH3 and PX domain...	2	
☐	DDA_Lib	sp|A2RRP1|NBAS_...	NBAS	NBAS subunit of NR...	6	
☐	DDA_Lib	sp|A3KMH1|VWA8_...	VWA8	von Willebrand facto...	6	
☐	DDA_Lib	sp|A3KN83|SBNO1_...	SBNO1	Protein strawberry n...	3	
☐	DDA_Lib	sp|A4D1E9|GTPBA_...	GTPBP10	GTP-binding protein ...	4	
☐	DDA_Lib	sp|A4D1P6|WDR91...	WDR91	WD repeat-containin...	3	

- If the user wants to sort the whole data, please use the global sorting function located on the top right corner of the "Proteins" panel. The user can select the column for sorting, then click the buttons for ascend or descend sort.

Selected ∨   A↓Z   Z↓A   ⑂⑂

Selected		...	Razor SpectralC..
Protein			
ProteinID		5	
EntryName		3	
Gene		0	
Length		2	
Organism		3	
ProteinDescription		9	

Total Spe...

3		3
9		9
4		3
2		2
7		7
6		6
3		3
4		4
3		3

1/5

How to move the pages?
To save the computing source and improve the visualization speed, there are only 1,000 proteins displayed in one page. The user could use the forward and backward arrows or type the page numbers in the text filed jumping to the desired page.

Total SpectralCo...	Unique Spectral...	Razor SpectralC..
3	3	
11	1	
4	0	
46	1	
2	0	
3	1	
2	2	
2	2	
2	2	
3	3	
3	3	
3	3	
9	1	

3   3/5

3. "Coverage" panel:
The sequences coverage of protein(s) selected in the "Proteins" panel will be shown here. Peptides in green background are uniquely identified or assigned to the current proteins

only. Peptides in grey background are shared with other proteins and not assigned to the current protein.

Coverage

sp|A0A2R8Y4L2|RA1L3_HUMAN

M R D P N T K R S R G F G F V T Y A T V E E V D A A M N A R P H K V D G
R V V E P K R A V S R E D S Q R P D A H L T V K K I F V G G I K E D T E
E H H L R D Y F E Q Y G K I E V I E I M T D R G S G K K R G F A F V T F
D D H D S V D K I V I Q K Y H T V N G H N C E V R K A L S K Q E M A S A
S S S Q R G R S G S G N F G G G R G G G F G G N D N F G R G G N F S G R
G G F G G S R G G G G Y G G S G D G Y N G F G N D G S N F G G G G S Y N
D F G N Y N N Q S S N F G P M K G G N F E G R S S G P H G G G G Q Y F A
K P R N Q G G Y G G S S S S S S Y G S G R R F

4. "PSMs" panel:

All PSMs identified in all proteins selected in the "Proteins" panel will be shown here. This information is from "psm.tsv" file. Same as the "Proteins" panel, the user can select the interested columns and sort PSMs globally using the functions located on the top right corner. If many proteins or all proteins are selected in the "Proteins" panels, the PSMs number will be huge. So same as the "Proteins" panel, only 1,000 PSMs will be shown on one page. And the user can change the page using the functions on the bottom right corner of the "PSMs" panel.

How to export annotated spectrum in batch?
If the user wants to export multiple annotated spectra in batch, please check the interested PSMs on the first "Selected" column or check the "Whole page" box to export all of them. Then click the button on the top left corner of the "PSMs" panel to specify the output path and the type and size of the image.

5. Annotated peptide spectrum match panel:
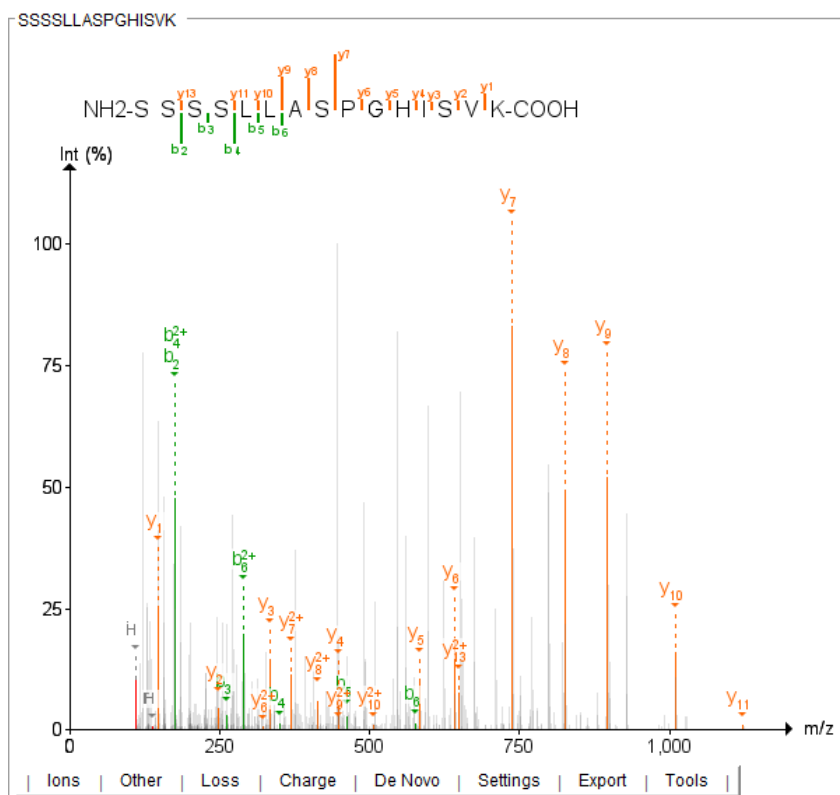
The selected PSM in the "PSMs" panel will be shown here. This panel includes the fragments panel on the top. The user can move it around by clicking and holding the mouse left key. It can also be zoomed in an out by using the mouse wheel.

The middle panel is the annotated spectrum. The m/z and relative intensity values will be shown if the user put the mouse on the annotated peak. Note that only annotated peaks will have m/z and intensity values shown unless the "Show All Peaks" option is checked in the Settings menu. The spectrum can also be zoomed in by clicking and holding the mouse left key. And it can be restored back to the defalt scale by clicking mouse right key.
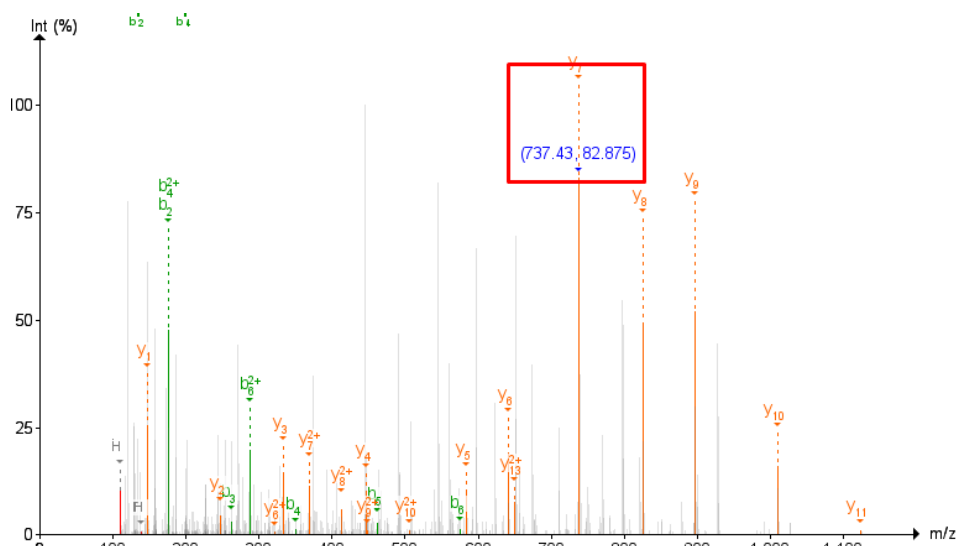
The bottom toolbox panel has several menus for settings that control the spectrum annotation settings.

| Ions | Other | Loss | Charge | De Novo | Settings | Export | Tools |

- Let's start with the "Ions" menu:
  This menu is designed for selecting fragment ion types shown in the spectrum.
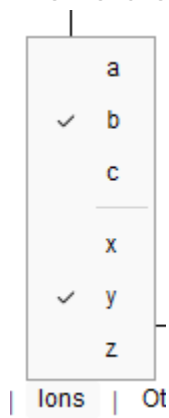
If multiple ions are selected, the fragment panel will show every ion in the following way.
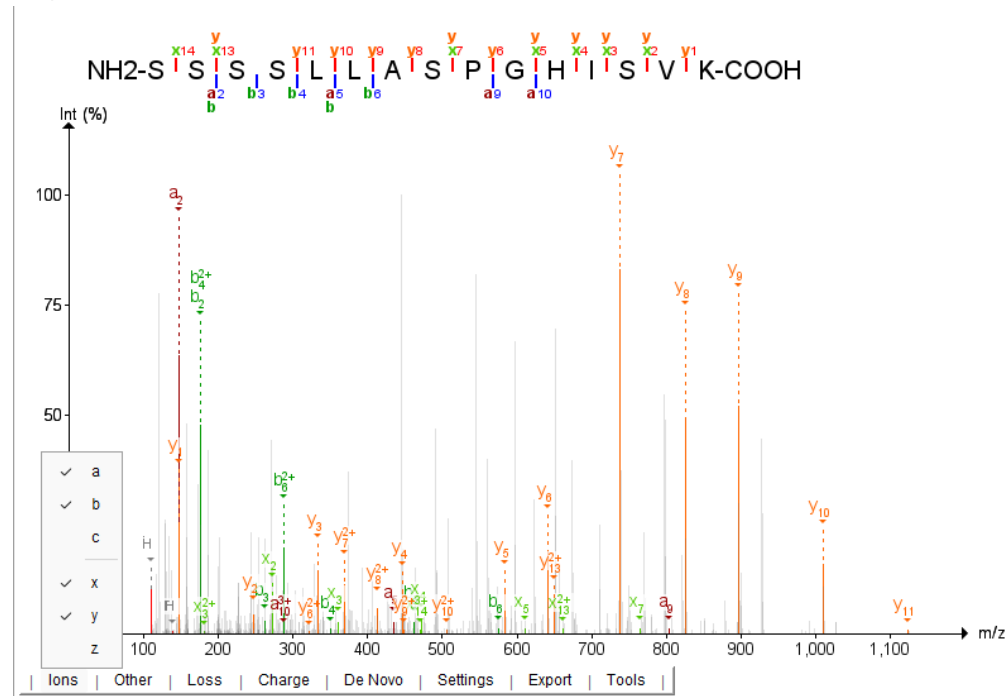
- The second menu "Other" includes all other ions the user wants to visualize.

If the "Glycan: B/Y" is clicked, several possible glycan fragment B/Y ions will be annotated in the spectrum, based on an internal list of common fragments. Note that the glycan fragments are NOT unique to the glycan identified in the spectrum and instead come from the entire list, so spurious matches (and misses) can occur. This does not necessarily mean that these ions were used in generating the search engine score.
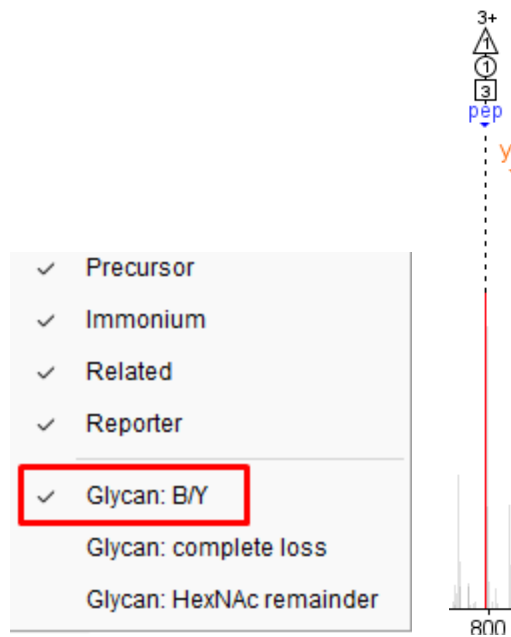
If the "Glycan: complete loss" is clicked, the glycosylation mass on "N" will be set as 0. If the "Glycan: HexNAc remainder" is clicked, the glycosylation mass on "N" will be set as the mass of HexNAc (203.0794), reflecting common fragmentation behavior of glycopeptides in collisional activation (CID, HCD) spectra.

If the O-pair mode is enabled during the FragPipe search, there is one function to show paired ETD spectrum if it is matched.
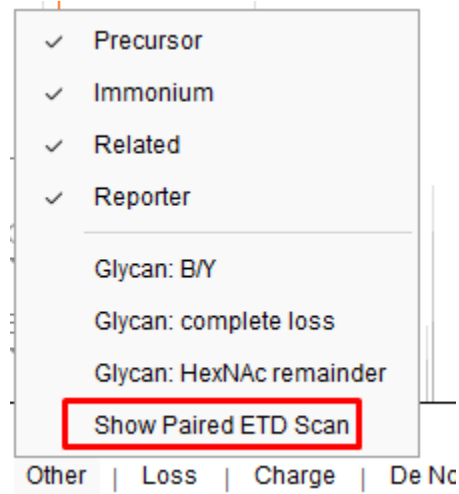
- The next menu "Loss" includes neutral losses. No loss is selected in default. If there is the phosphorylation in the search, the H3PO4 will be added to the menu automatically.
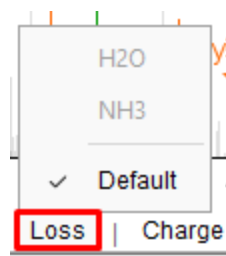
- The next menu "Charge" shows the charge states of fragments the user wants to annotate. The highest charge state is up to the precursor charge.
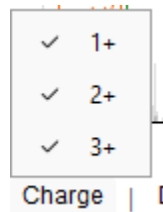
- The next menu "De Novo" gives the user one option to annotate the spectrum with amino acids based on the identification result.
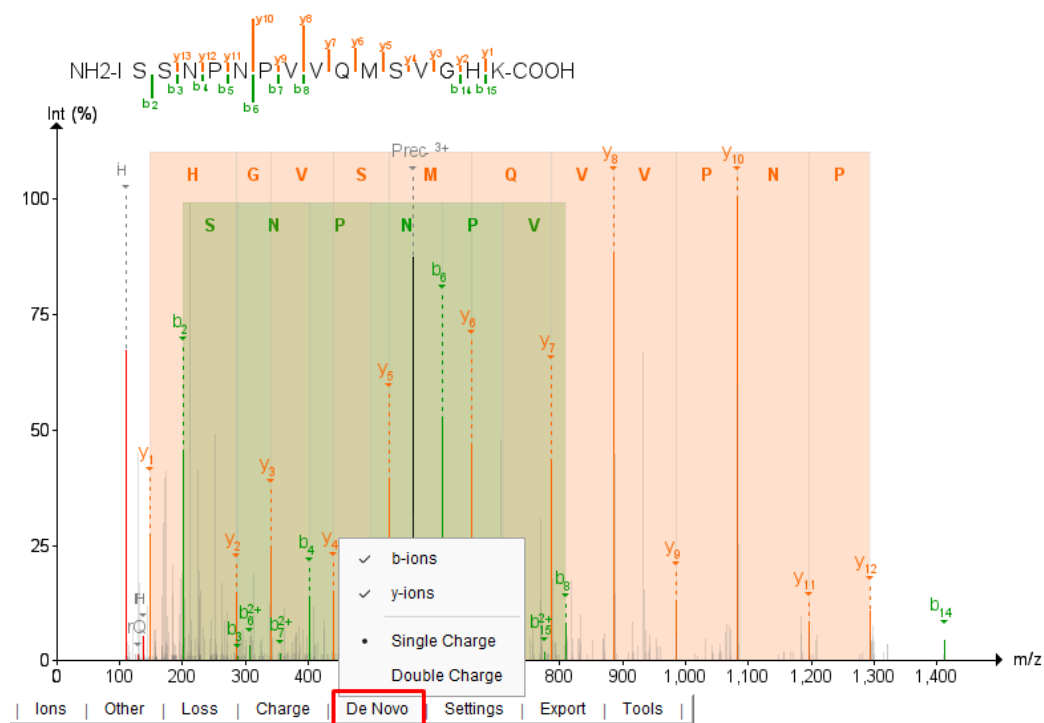
- The next menu "Setting" includes some detail configurations. (**BE CAREFULL, ALL SETTINGS ARE GLOBAL**)

The first item is "Show all Peaks". In default, FragPipe-PDV only shows the annotated peaks in color and the unannotated peaks will be grey out as the background. After clicking the "Show All Peaks", every peak will be same, and the use can see every peak's detail information using mouse.
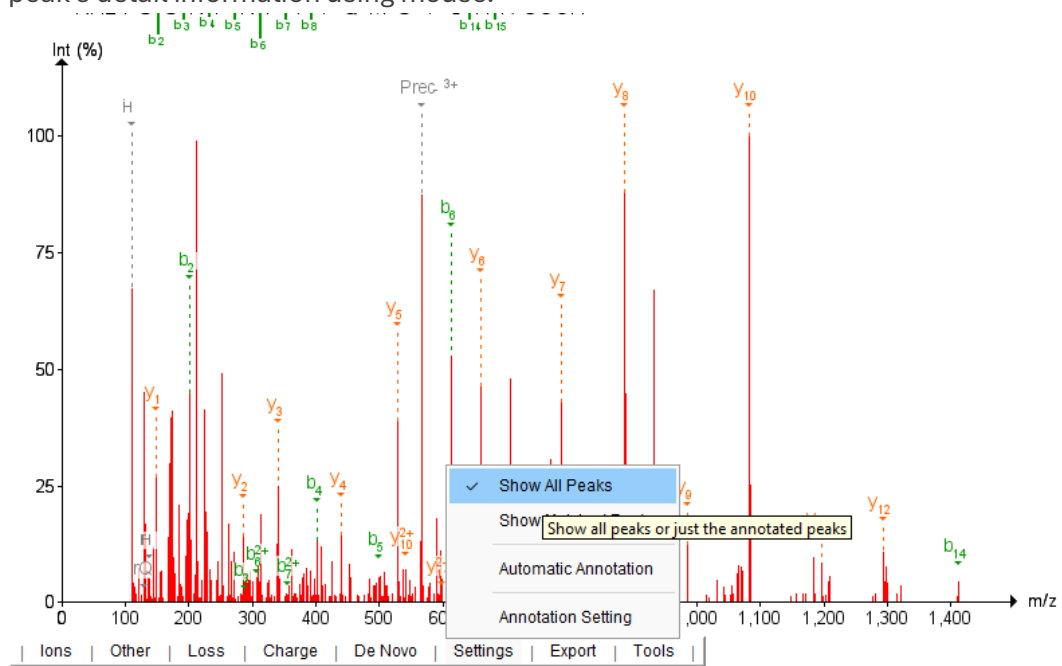
Some time there is a background peak with high intensity (e.x. precursor peak) as follows:
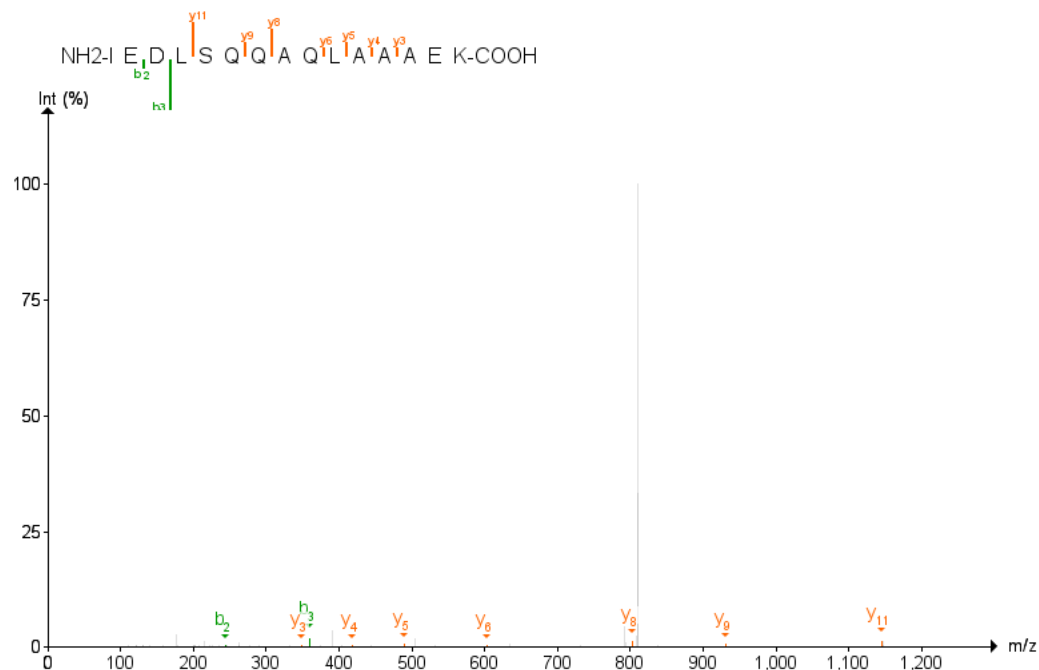
And we can want to see the annotated peaks only. Then the user can check the "Show Matched Peaks". All background peaks without annotations will be removed.
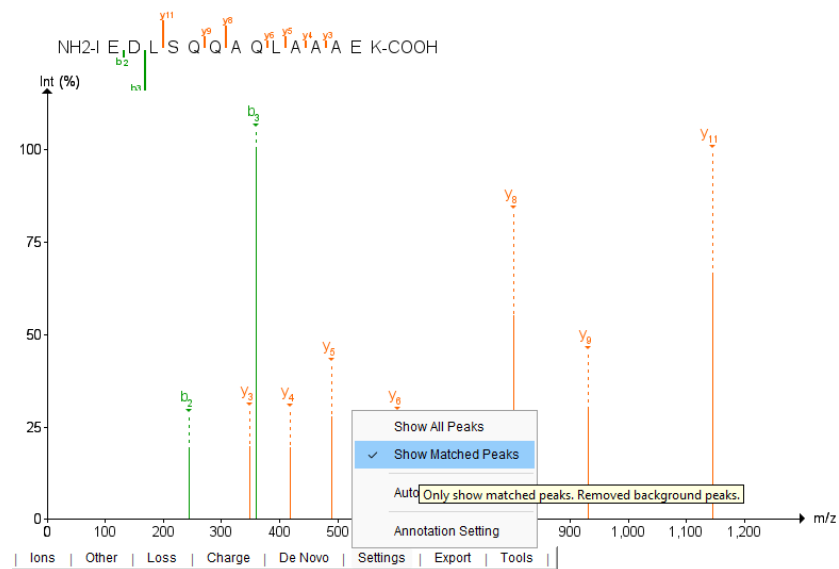
Just ignore the "Automatic annotation"

After clicking the "Annotation Setting", a new panel will pop up:

PDV - Annotation settings ✕

Annotation Colors

	Ion	
1	b	🟩
2	b-H2O	🟦
3	b-H2O-NH3	⬜
4	b-C3H9N-H2O	⬜
5	b-H2O-HPO3	⬜
6	b-CH4OS-H2O	⬜
7	b-H2O-H3PO4	⬜
8	b-NH3	🟪
9	b-C3H9N-NH3	⬜
10	b-HPO3-NH3	⬜
11	b-CH4OS-NH3	⬜
12	b-H3PO4-NH3	⬜

Peak Settings

Annotated Peak Color 🟥   Annotated Peak Width 1   Glycan Structure ☐

Background Peak Color ⬜   Background Peak Width 1

Annotation Level

Annotation Limit 0   Fragment m/z Tolerance: 20.0 ppm ⌄

OK

The top table is used to change the color of the annotated peaks. Click the color in the table and chose the new color:
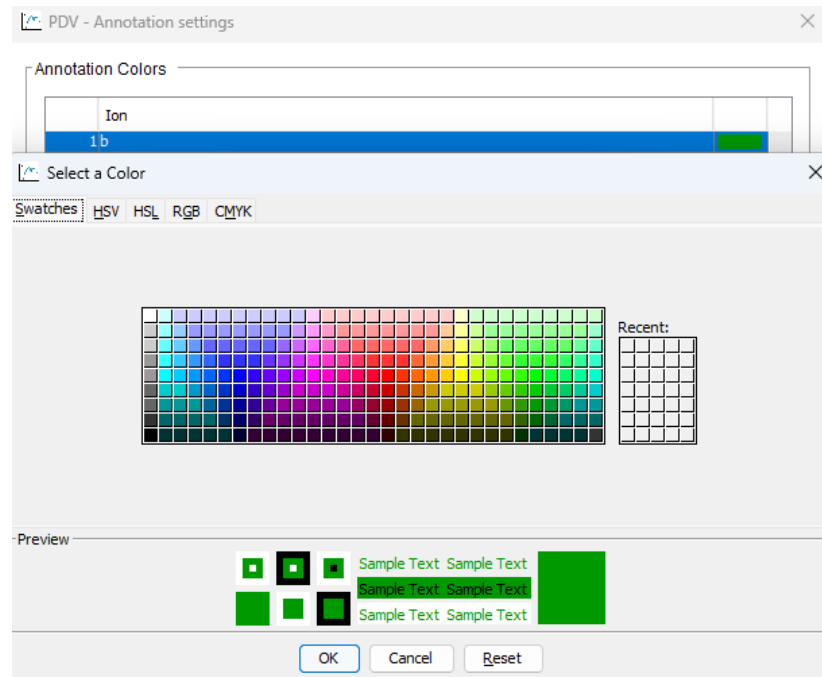
The "Peak Settings" in the middle includes the "Annotated Peak Color" for regular annotated peaks; the "Background Peak Color" for background peaks; "Annotated Peak Width" for annotated peak line width; the "Background Peak Width" for background peaks line width; a fake glycan structure will be showed if the 'Glycan Structure" is checked.
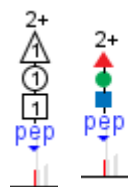
(**BE CARFUL, THIS STRUCTURE IS FOR VISUALIZATION ONLY AND IS NOT INTENDED TO INDICATE THE TRUE GLYCAN STRUCTURE**)

The "Annotation Level" on the bottom includes "Annotation Limit" for filtering low intensity peaks during annotation. The number here means peaks below this relative intensity will not be annotated. For example, a setting of 0.01 means only peaks above 1% relative intensity will be annotated, all others will be filtered out. Set to 0 to annotate all peaks. "Fragment m/z Tolerance" is the tolerance for mathcing peaks.

- The next menu "Export" includes export options for current spectrum. It could be an image or a mgf file:
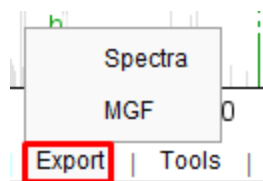
If the user wants to export current spectrum as an image, a pop-up panel will be shown after clicking on the "Spectra".
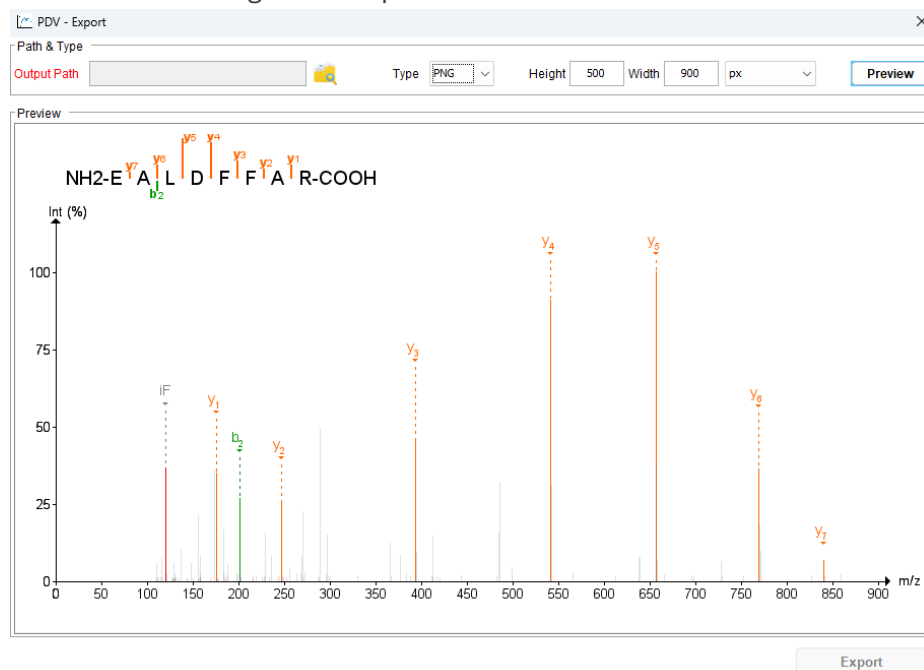
First specify the "Output Path"; second select the image type including PNG, TIFF, and PDF; third specify the size and size unit. If the user updates the size, please click "Preview" button to update the spectrum. The output image will be the same as what is displayed here.

- The last menu is "Tools". We add many useful functions here. The default one is "Show spectrum" which is the normal viewer.
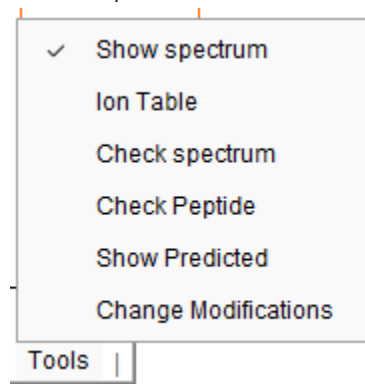
The user can select "Ion Table" to show the annotation ions table:

EALDFFAR

	b		b++	AA	y		y++	
1	130.05		65.53	E				8
2	201.09		101.05	A	839.44		420.22	7
3	314.17		157.59	L	768.40		384.71	6
4	429.20		215.10	D	655.32		328.16	5
5	576.27		288.64	F	540.29		270.65	4
6	723.33		362.17	F	393.22		197.12	3
7	794.37		397.69	A	246.16		123.58	2
8				R	175.12		88.06	1

If the user wants to check the annotation of current peptide with another spectrum, we can click "Check spectrum" function. A new button "Add" with "Add Spectrum" will be showed. The user can import a spectrum file in mgf format. Then FragPipe-PDV will compare two PSMs in a mirror spectrum format.

If the user wants to check the annotation of current spectrum with another peptide or different modifications, we can click "Check peptide" button. A new button "Add" with "Add Peptide" will be showed.

After clicking "Add Peptide" button, a new panel will be displayed. The default peptide displayed is the same as the original.

The user can change the sequence in the text filed.

The user can also change or add new modifications on the amino acid by clicking the amino acid on the second panel:

The amino acid with modification will be labeled as blue:

Finally, click the "Add" button to add the comparison.
A mirror spectrum will be displayed:
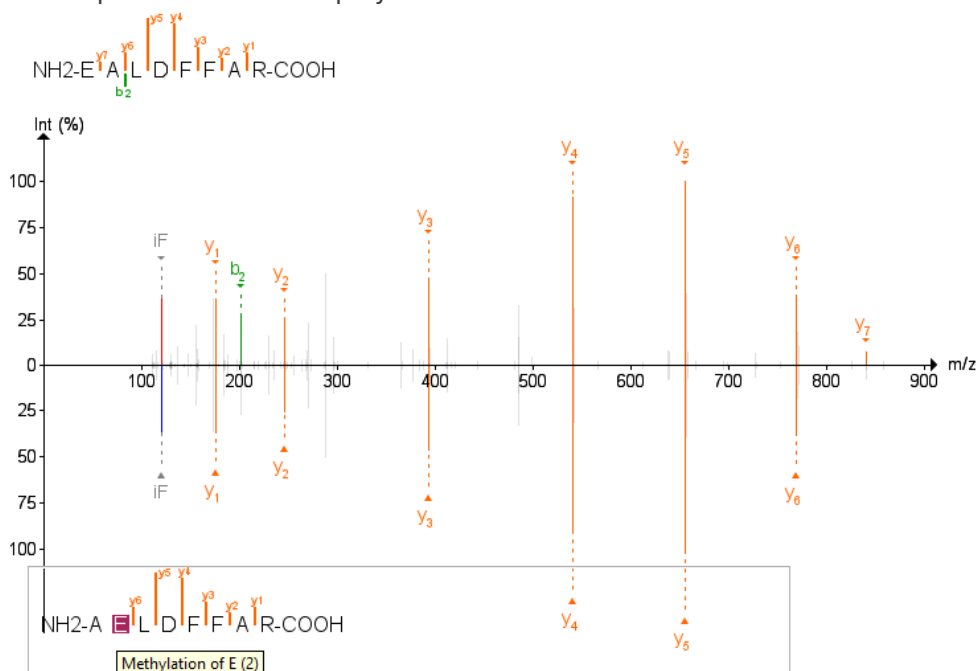
Since MSBooster in FragPipe is applied in default, FragPipe-PDV can get the predicted spectrum from MSBooster. So after clicking "Show Predicted", the predicted spectrum will be displayed in mirror spectrum format. The "Unweighted Spectral Entropy" is also showed here. (If MSBooster is disable during search, this function will be disabled.)

The last function is "Change Modifications". After clicking it, a new panel will pop up. In this function, the user can 1. Modify the known modification mass and 2. Assign the unlocated "Delta mass" to any amino acid.
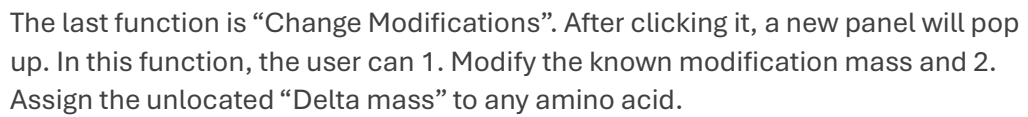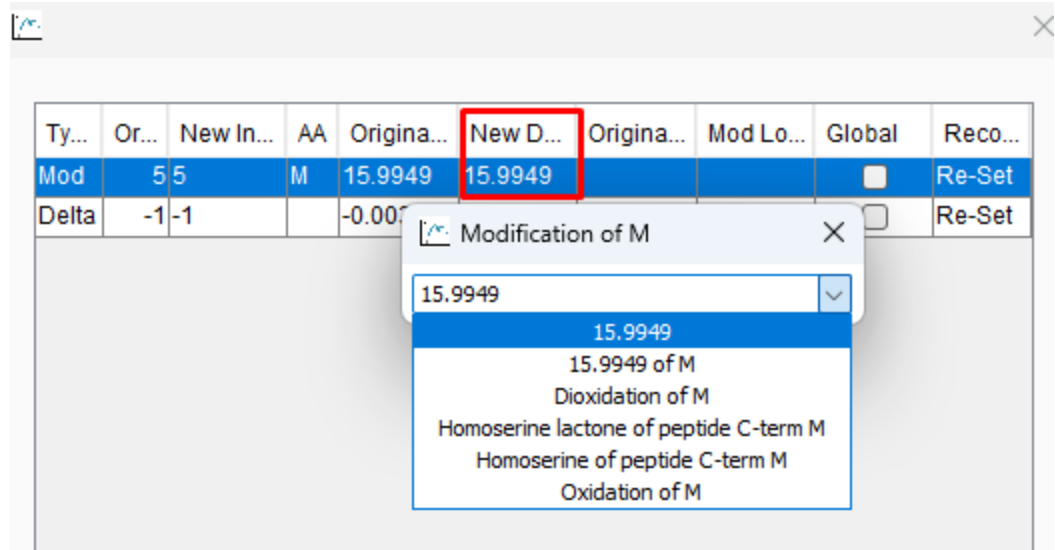

Ty...	Or...	New In...	AA	Origina...	New D...	Origina...	Mod Lo...	Global	Reco...
Mod	5	5	M	15.9949	15.9949			☐	Re-Set
Delta	-1	-1		-0.0031	-0.0031			☑	Re-Set

OK

For example, there is an oxidation on M at 5th amino acid. If I want to change this modification to another modification or any mass, I can click the "New Defined" column and assign a new modification or type any mass.

There is a "Delta mass" without location reported by FragPipe which is normal in mass offset or open search. I can click the "New Index" column and assign one location for it.
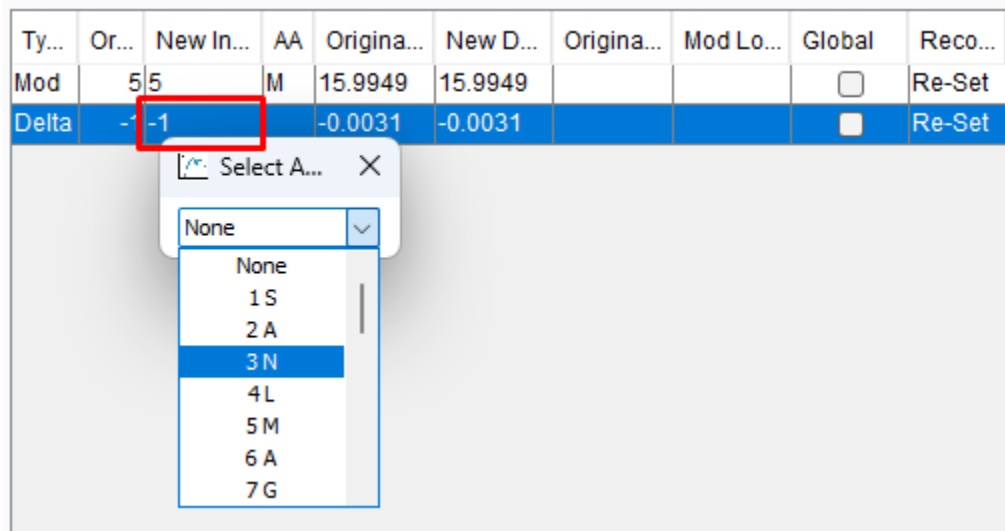
Troubleshooting:

If FragPipe-PDV freezes, please try to close the FragPipe GUI first. Some unseen conflicts occur accidentally between the two GUIs.
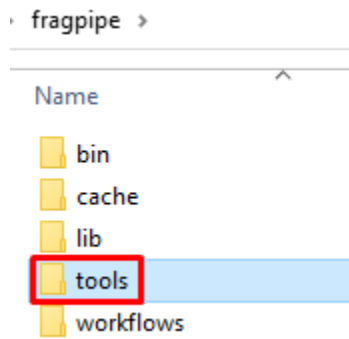
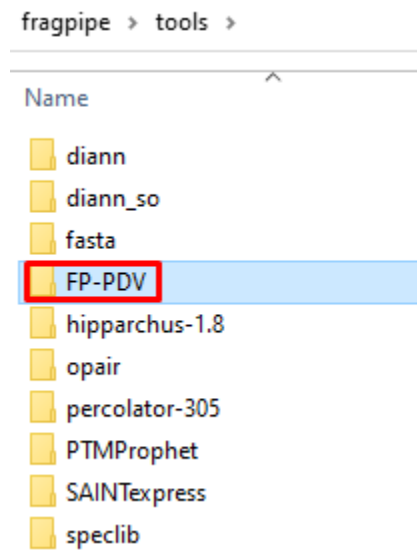For other issues, feel free to submit an issue through FragPipe GitHub: https://github.com/Nesvilab/FragPipe/issues

Update FragPipe-PDV in FragPipe manually:

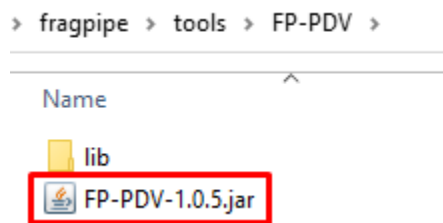The latest FragPipe-PDV packages are free to download at https://github.com/Nesvilab/FragPipe-PDV/releases.

(1) Open the tools in your FragPipe folder.

(2) Open the FP-PDV in the tools folder.

(3) Copy the latest FragPipe-PDV package and replace the old one in the FP-PDV folder.

How to run FragPipe-PDV in CMD:

java -jar FP-PDV-1.0.5.jar your_result_folder threads_num