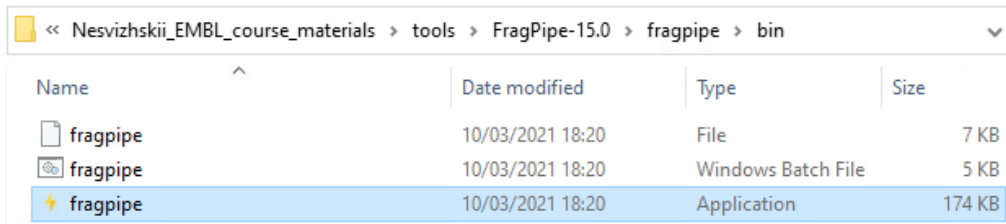


Open search strategy for PTM analysis

Setup FragPipe

From the Desktop, open the Nesvizhskii_EMBL_course_materials folder, and open the tools subfolder. Unzip both the FragPipe-15.0 and MSFragger-3.2 files: double click to view compressed folder tools, then select 'Extract all' (you can optionally remove FragPipe-15.0 from the end of the path so it reads C:\Users\Public\Desktop\Nesvizhskii_EMBL_course_materials\tools), then click 'Extract'. Repeat the unzipping procedure for the MSFragger3.2 file. From FragPipe-15.0\fragpipe, open the bin subfolder and double click the file with the lightning bolt icon to launch the program.



Name	Date modified	Type	Size
fragpipe	10/03/2021 18:20	File	7 KB
fragpipe	10/03/2021 18:20	Windows Batch File	5 KB
fragpipe	10/03/2021 18:20	Application	174 KB

FFPE artifact discovery

One common use for open searches is finding experimental artifacts that can be included in subsequent closed searches to increase proteome coverage. Many sample preparation methods can modify peptides and reduce the likelihood of recovering them in a typical search. One such protocol is formalin-fixed paraffin-embedding (FFPE), a widely used tissue preservation method. There are a few different modification palettes that have been suggested in the literature (Metz et al., *J. Biol. Chem.*, 2004; Hood et al., *Mol. Cell. Proteom.*, 2005; Zhang et al., *Proteomics*, 2015) and we are interested in knowing which of these modifications, if any, are most relevant to our data.

Sample (from SCIEX TripleTOF .raw, converted to .mzML):

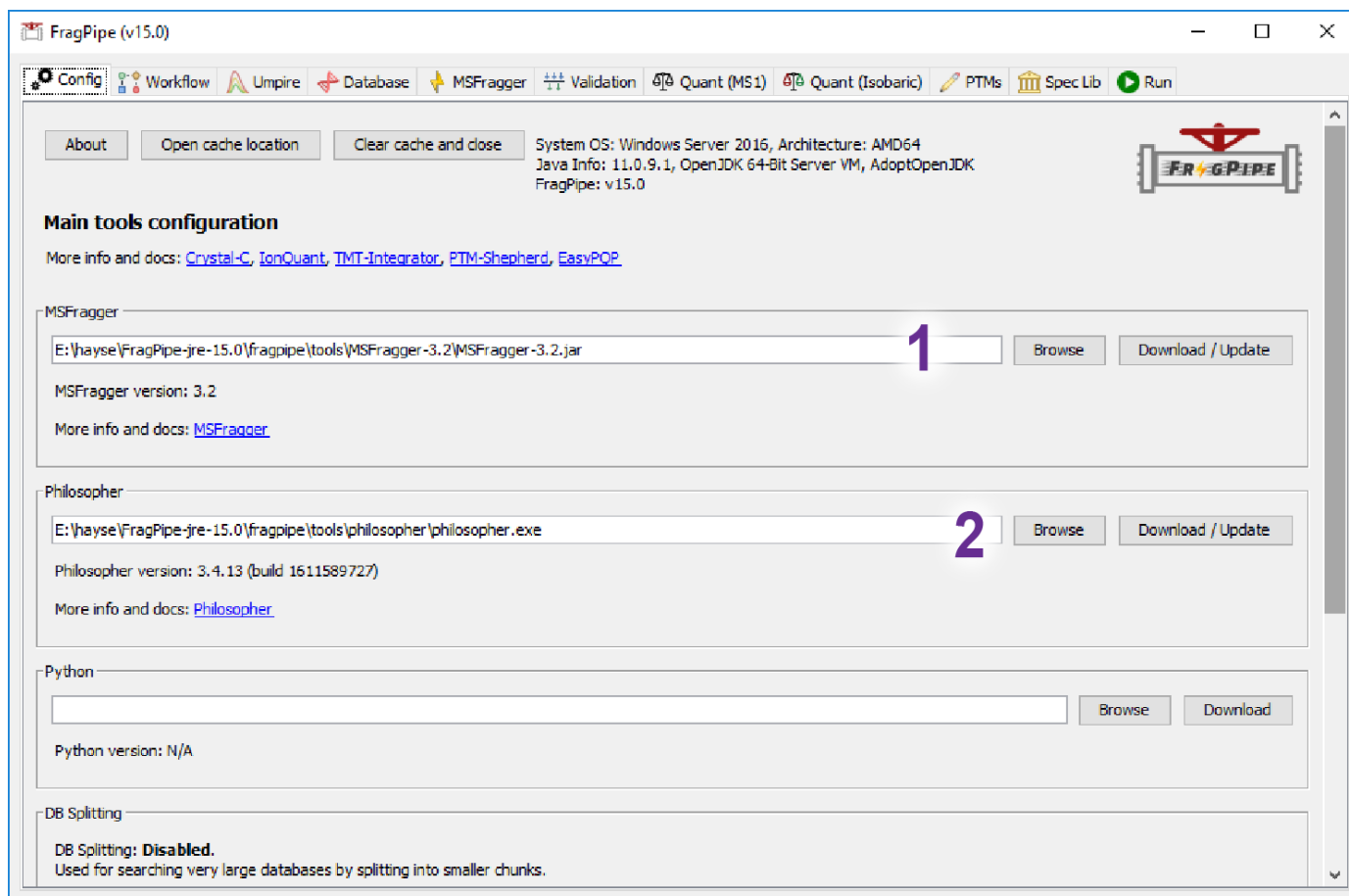
FFPE preserved amyloid deposits in eye tissues (2014-03-14_-_NSN_-_38B_-_2.mzML)

(Publication: Nielsen, Nadia Sukusu, et al. "Insight into the protein composition of immunoglobulin light chain deposits of eyelid, orbital and conjunctival amyloidosis." *Journal of proteomics & bioinformatics* (2014).)

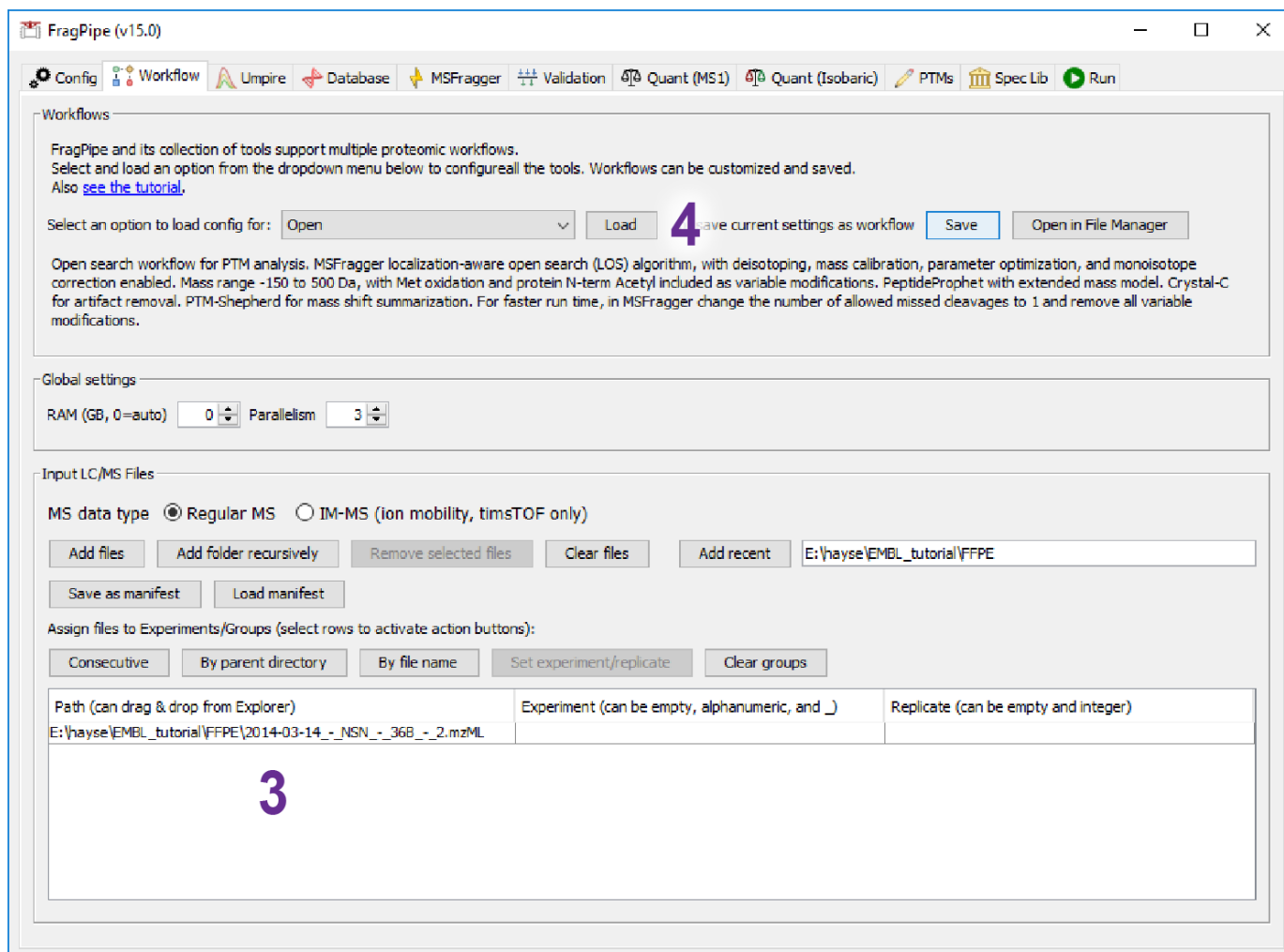
Open search with FragPipe

1. FragPipe starts on the Config tab where we configure the pipeline's dependencies. Make sure FragPipe can find the MSFragger executable for database searching, the path should be C:\Users\Public\Desktop\Nesvizhskii_EMBL_course_materials\tools\MSFragger-3.2\MSFragger-3.2
MSFragger can also be downloaded using the 'Download / Update' button if needed.
2. Philosopher is needed to filter the search results and generate reports, make sure FragPipe can find the philosopher executable, the path should be C:\Users\Public\Desktop\Nesvizhskii_EMBL_course_materials\tools\philosopher-3.4.13

We can also download Philosopher with the 'Download / Update' button if needed.



3. On the Workflow tab, drag and drop the file 2014-03-14_-_NSN_-_36B_-_2.mzML file from the FFPE folder. Since we are only analyzing a single file, we don't need to provide experiment or replicate labels.
4. Fragpipe includes built-in workflows for many common analyses. It is recommended to use the defaults workflows as a starting point for any custom analyses. Since we will be doing an open search, select the 'Open' workflow from the dropdown menu at the top of the page. Click 'Load' to configure FragPipe to run a complete open search workflow.



5. On the MSFragger tab, have a look at the search parameters we will use. Note that a wide precursor mass tolerance is set, from -150 Da to +500 Da. In the Protein Digestion section, change the number of allowed missed cleavages from 2 to 1.

FragPipe (v15.0)

Config Workflow Umpire Database **MSFragger** Validation Quant (MS1) Quant (Isobaric) PTMs Spec Lib Run

☒ Run MSFragger

Save Config Load : Custom MSFragger parameter file from disk

Common Options (Advanced Options are at the end of the page)

Peak Matching

Precursor mass tolerance Da -150 - 500 Fragment mass tolerance PPM 20

Calibration and Optimization Mass calibration, parameter optimization Isotope error 0 Data type DDA

Protein Digestion

Load rules stricttrypsin Enzyme name stricttrypsin Cut after KR But not before

Cleavage ENZYMATIC Missed cleavages 2 ☒ Clip N-term M

Peptide length 7 - 50 Peptide mass range 500 - 5,000 Split database 1

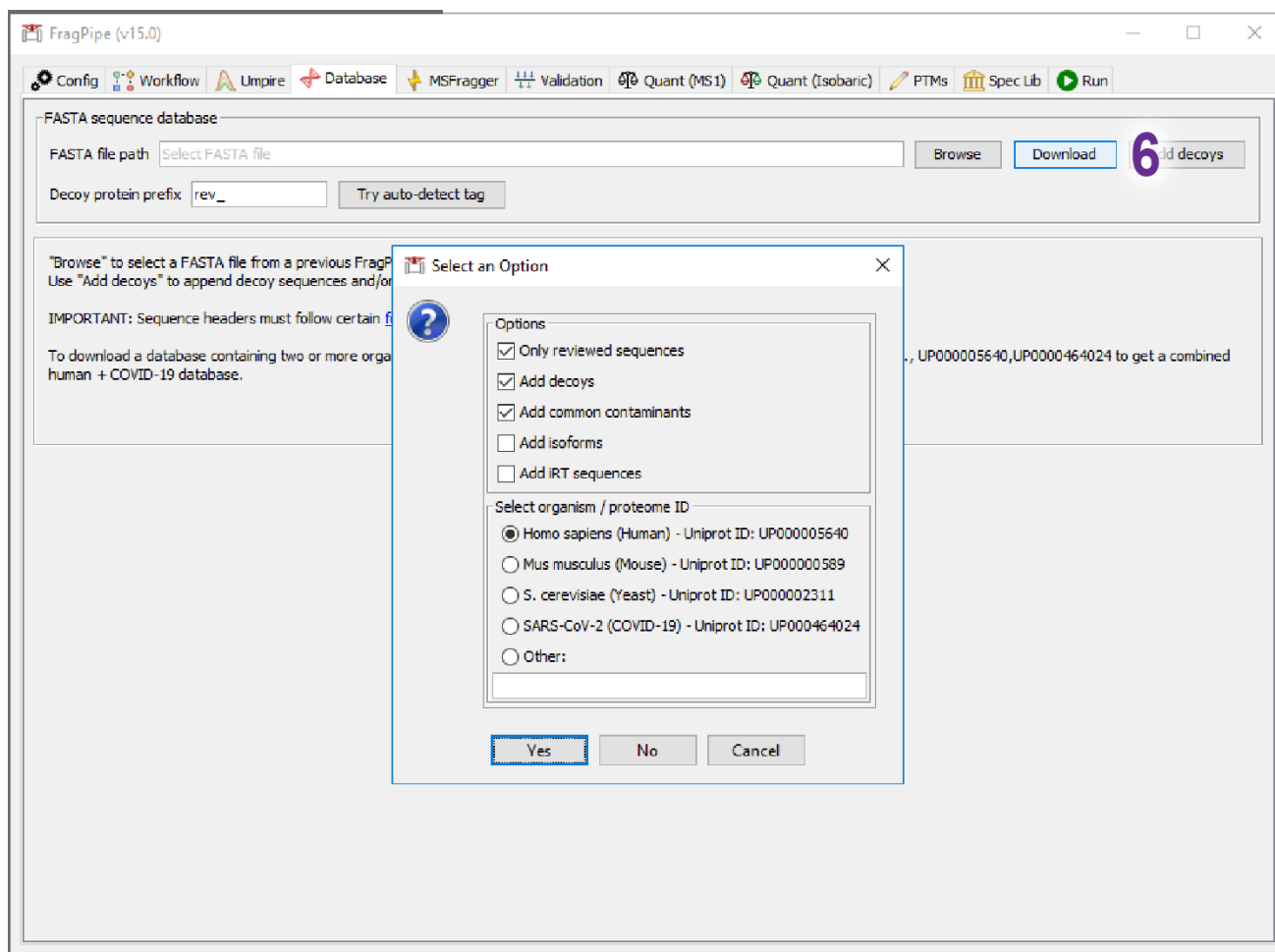
Modifications

Variable modifications

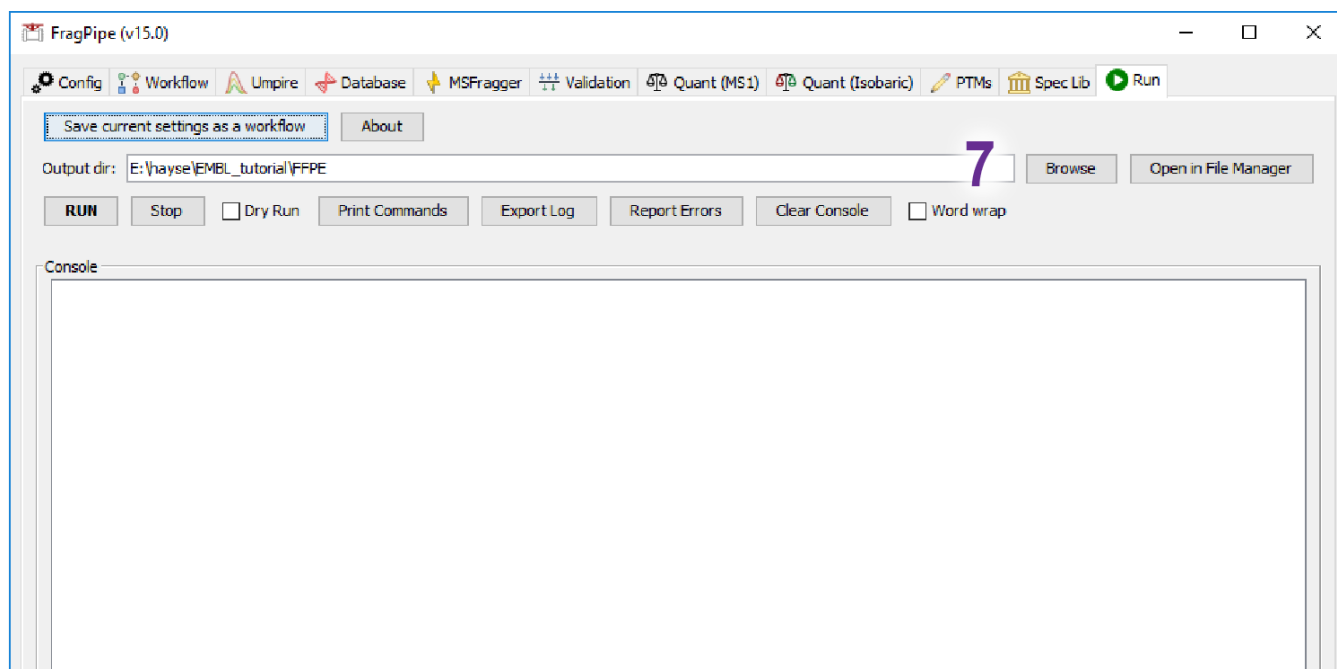
Max variable mods on a peptide 3 Max combinations 5,000

Enabled	Site (editable)	Mass Delta (edita...	Max occurrences ...
<input checked="" type="checkbox"/>	M	15.9949	3
<input checked="" type="checkbox"/>	[^	42.0106	1
<input type="checkbox"/>	STY	79.96633	3
<input type="checkbox"/>	nQnC	-17.0265	1

- Now we need to select a protein sequence database. You can choose the file 2021-04-14-decoys-reviewed-contam-UP000005640.fas in the course materials folder, or you can download one using FragPipe. Downloading is easy, so we could also choose to download one at this point. On the Database tab, click the 'Download' button. Follow the prompts to use the default settings (reviewed human sequences with common contaminants).



7. All other tabs have already been configured for a basic open search analysis, so skip to the Run tab to set the output location (a new folder called 'my_FFPE_results'), then click 'RUN' and wait for the results.



Examine the open search results

In the my_FFPE_results folder, you will see PTM-Shepherd output files ('global.profile' and 'global.modsummary') that help interpret all the mass shifts identified from the open search.

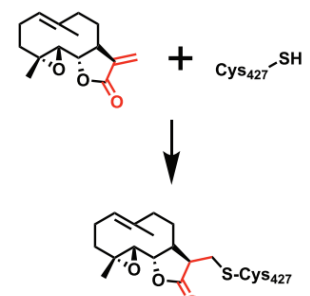
To open the global.profile report, right click and 'Open With', find C:\Program Files (x86)\Microsoft Office\Office16\EXCEL.

In the global.profile file:

- Deamidation is commonly included in proteomics searches. If being more abundant than deamidation were the bar for inclusion in an FFPE PTM-palette, how many modifications would be included? (Hint: peak_signal is a measure of relative peak quality. It is low if we have a hard time attributing all of the PSMs in a particular range to the closest mass shift, such as would be the case in a low abundance modification if it were immediately adjacent to an abundant one.)
- Modification localization tells us which residues our PTMs are found on. PTM-Shepherd produces an N-terminal localization rate as a percentage of PSMs that could be localized to the N-terminus and residue-specific localization scores. Residue-specific localization scores are broken into three parts: the amino acid, the enrichment score (roughly equivalent to an odds ratio), and the estimated PSMs it was found on. We searched for oxidation of M as a variable modification. Which other residues is oxidation found on?
- PTM-Shepherd uses a database called UniMod to annotate mass shifts. The mass shift of +32 Da mapped to a dihydroxy event. Do you think this is correct? (Hint: Compare how confident PTM-Shepherd was in localizing dihydroxy and single oxidation events to proline. You can compare enrichment scores and calculate the % of modifications that were successfully localized by dividing the 'Localized_PSMs' column by the 'PSMs' columns.) What could cause a mass shift to be less confidently localized?
- PTM-Shepherd was unable to identify one mass shift of +196 Da. Do you have any idea what this is? (Hint: We don't know what it is.) What might we gain from including an abundant, unknown mass shift in a subsequent search?
- For the mass shift of +28.03 Da, UniMod found three separate explanations for it. Go to unimod.org and see if you think we can rule any of them out based on their annotations.
- If the +32 mass shifts were attributable to two separate +16 modifications on the same peptide and they result in less confident localization, why would the +28.03 Da di-methylation peak appear to be more confidently localized than its single methylation counterpart?
- Write down your best guess for optimal parameters for a second pass search. Make sure to include the mass shift, the residue, and the number of occurrences per peptide.
- Bonus: Formylation is one of the mass shifts we might expect to be highly abundant given the nature of FFPE preservation. It looks like PTM-Shepherd was able to do some partial localization based on the percentage of PSMs that were localized, but there are no amino acids that are reported as being enriched. It also has a high modified-unmodified spectral similarity score, indicative of labile modifications. Does the formylation mass shift overlap with any other modifications? What metrics could we use to distinguish them? If formylation were more abundant in this dataset and we wanted to include it in another search, what do you think the best strategy for including a labile, ambiguously localized modification is?

Parthenolide labeling exploration

This chemical is frequently used by cytoskeleton researchers to inhibit tubulin detyrosinases, but it is suspected to hit more than just those enzymes. For



example, the figure on the right shows the mechanism of parthenolide covalently labeling Cys427 of FAK1 (Berdan et al., *Cell chemical biology*, 2019). Labeled peptides should have a +248.14125 Da mass shift, but we want to know what residues it modifies. This is a case for mass offset search, where the mass is known but we want to allow modification of any amino acid.

Samples (from Q Exactive HF .raw, converted to .mzML):

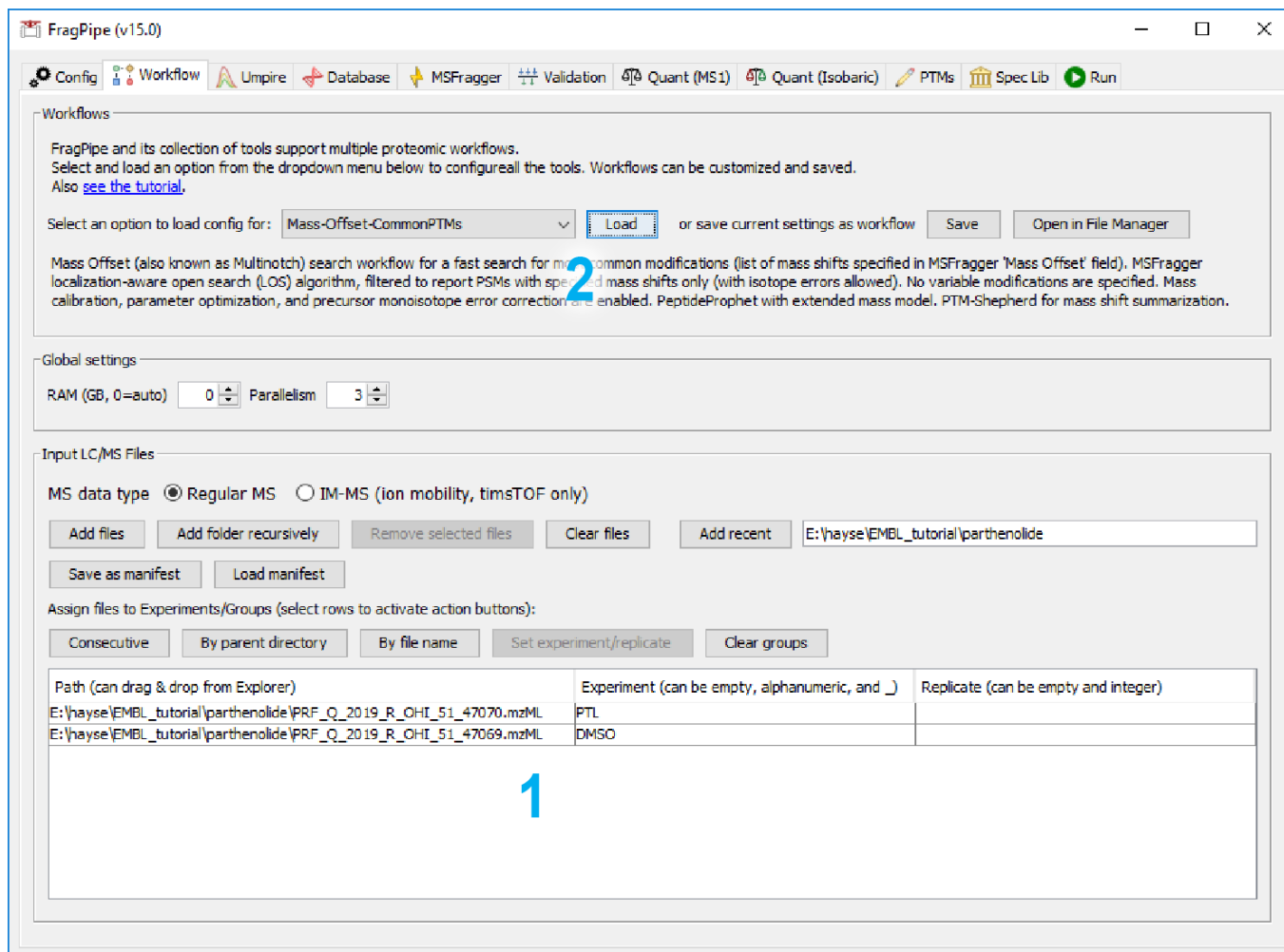
Purified tubulin treated with DMSO/control (PRF_Q_2019_R_OHI_51_47069.mzML)

Purified tubulin treated with 20 μ M parthenolide (PRF_Q_2019_R_OHI_51_47070.mzML)

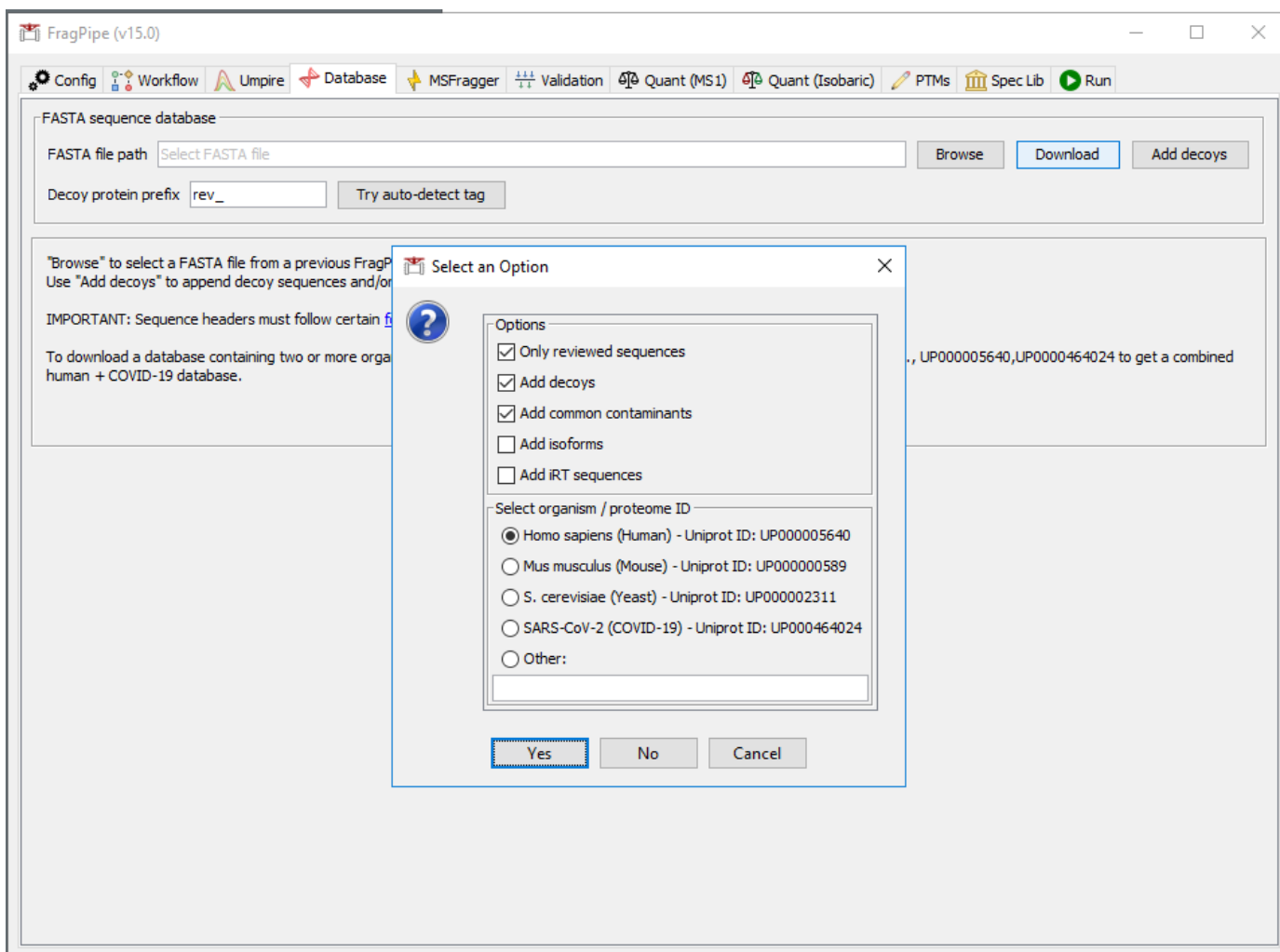
(Publication: Hotta, Takashi, et al. "Parthenolide Destabilizes Microtubules by Covalently Modifying Tubulin." *Current Biology* 31.4 (2021): 900-907.)

Mass offset search with FragPipe

1. Load the files (drag and drop or browse), specify experiments: "DMSO" for the file ending 70, "PTL" for the one ending 69
2. Load the Mass-Offset-CommonPTMs workflow



3. On the Database tab, ensure a database of reviewed human sequences is chosen (you can use the one from the previous analysis) or download one



4. In the MSFragger tab, check the first box in the Variable modifications section to set methionine oxidation as a variable modification. Also make cysteine carbamidomethylation a variable modification instead of a fixed mod: add C +57.02146 to variable modifications

FragPipe (v15.0)

Config Workflow Umpire Database MSFragger Validation Quant (MS1) Quant (Isobaric) PTMs Spec Lib Run

☒ Run MSFragger

Save Config Load : Custom MSFragger parameter file from disk

Common Options (Advanced Options are at the end of the page)

Peak Matching

Precursor mass tolerance PPM -10 - 10 Fragment mass tolerance PPM 20

Calibration and Optimization Mass calibration, parameter optimization Isotope error 0/1 Data type DDA

Protein Digestion

Load rules stricttrypsin Enzyme name stricttrypsin Cut after KR But not before

Cleavage ENZYMATIc Missed cleavages 2 ☒ Clip N-term M

Peptide length 7 - 50 Peptide mass range 500 - 5,000 Split database 1

Modifications

Variable modifications

Max variable mods on a peptide 3 Max combinations 5,000

Enabled	Site (editable)	Mass Delta (edita...	Max occurrences ...
<input checked="" type="checkbox"/>	M	15.9949	3
<input type="checkbox"/>	[^	42.0106	1
<input checked="" type="checkbox"/>	C	57.02146	3
<input type="checkbox"/>	nQnC	-17.0265	1
<input type="checkbox"/>	nE	-18.0106	1
<input type="checkbox"/>	K	4.025107	2
<input type="checkbox"/>	R	6.020129	2
<input type="checkbox"/>	K	8.014199	2

4

5. Then scroll down to set fixed Mass Delta to 0 for cysteine

6. In Mass Offsets, replace the list of mass offsets with just 0/248.14125 for the parthenolide adduct

FragPipe (v15.0)

Config Workflow Umpire Database MSFragger Validation Quant (MS1) Quant (Isobaric) PTMs Spec Lib Run

Fixed modifications

Enabled	Site	Mass Delta (editable)
<input checked="" type="checkbox"/>	C-Term Peptide	0
<input checked="" type="checkbox"/>	N-Term Peptide	0
<input checked="" type="checkbox"/>	C-Term Protein	0
<input checked="" type="checkbox"/>	N-Term Protein	0
<input checked="" type="checkbox"/>	G (glycine)	0
<input checked="" type="checkbox"/>	A (alanine)	0
<input checked="" type="checkbox"/>	S (serine)	0
<input checked="" type="checkbox"/>	P (proline)	0
<input checked="" type="checkbox"/>	V (valine)	0
<input checked="" type="checkbox"/>	T (threonine)	0
<input checked="" type="checkbox"/>	C (cysteine)	0

5

Advanced Options

Mass Offsets

0/248.14125 6

Restrict delta mass to all

Glyco/Labile mode

Labile Modifications Search mode off Diagnostic Ion Minimum Intensity 0

Y Ion Masses 0 203.07937 406.15874 568.21156 730.26438 892.3172 349.137279

Diagnostic Fragment Masses 204.085646 186.076086 168.065526 366.139466 144.0656 138.055 126.055 163.060096 512.197375 292.1026925 274.0921325 657.2349 243.026426 405.079246 485.045576 308.09761

Spectral Processing

Precursor mass mode corrected

Min peaks 15 Use top N peaks 150 Min ratio 0.01

Clear m/z range 0 - 0

- On the PTMs tab, remove 'Failed_Carbamidomethylation:-57.021464' from the Custom mass shifts box.
- On the Run tab, make a new folder for the output files ('my_parthenolide_results'), then click 'RUN' and wait for the analysis to finish.

Examine the mass offset search results

In the results folder, you will see subfolders containing individual results for DMSO and PTL, plus PTM-Shepherd output files that summarize the mass shifts from both experimental conditions together. In the PTM-Shepherd 'global.profile' report:

- Each mass shift is listed in the peak_apex column. Look at the number of PSMs found for each mass shift between the two samples (DMSO and PTL).
- Which amino acid is primarily labeled with the +248 (first isotope, +249) mass shift?
- Is this the only amino acid enriched with this modification?

RNA crosslinking analysis

[XRNAX](#) is a method for “unbiased” purification of protein-crosslinked RNA, which can be used to help probe and understand protein-RNA interactions.

Sample (from QExactive HF .raw, converted to .mzML):

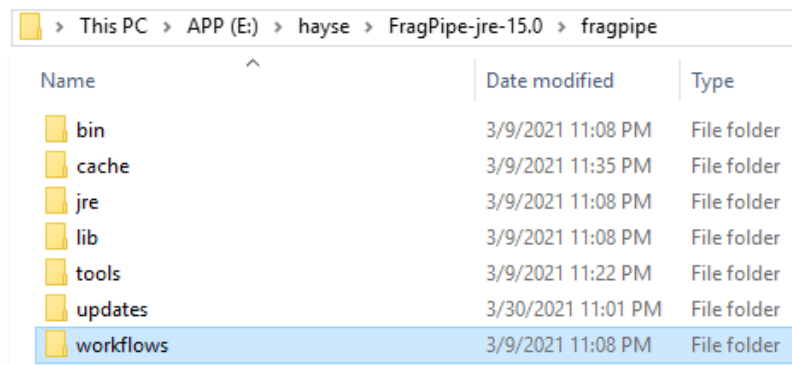
RNA-crosslinked tryptic peptides from MCF7 cell line (20160530_QE1_JT_XRNAXpep_Trp_hydro.mzML)

(Publication: Trendel, Jakob, et al. "The human RNA-binding proteome and its dynamics during translational arrest." *Cell* 176.1-2 (2019): 391-403.)

Custom workflow use with FragPipe

1. For this analysis, we will need to load a custom workflow. This workflow searches the database with mass offsets from the crosslinked RNA fragments. One can perform a fully open search, as it was done in the paper (if so, increase the mass window to (-150 Da 1000 Da)). However, mass offset searches are faster and they are also more sensitive than performing a full open search (but the drawback is that you can only find peptides with mass shifts you specified).

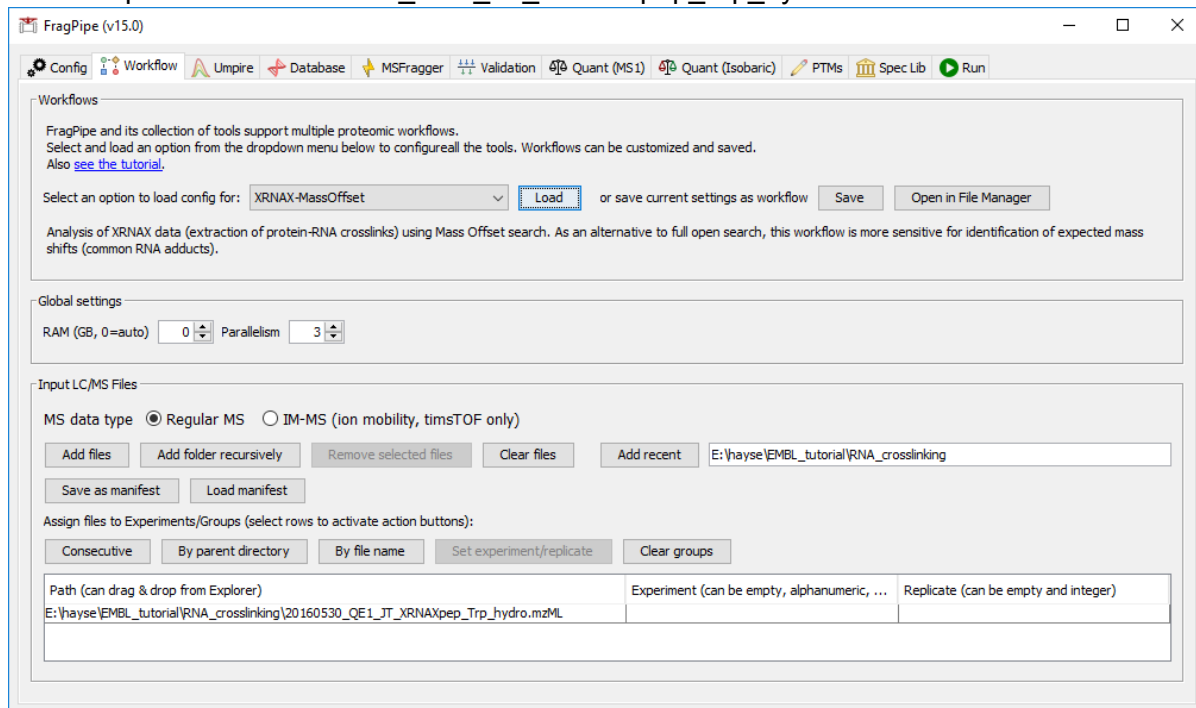
Find the XRNAX-MassOffset.workflow file in the RNA_crosslinking folder. Then in the fragpipe directory, locate the workflows subfolder, and copy the XRNAX-MassOffset.workflow file into this folder.



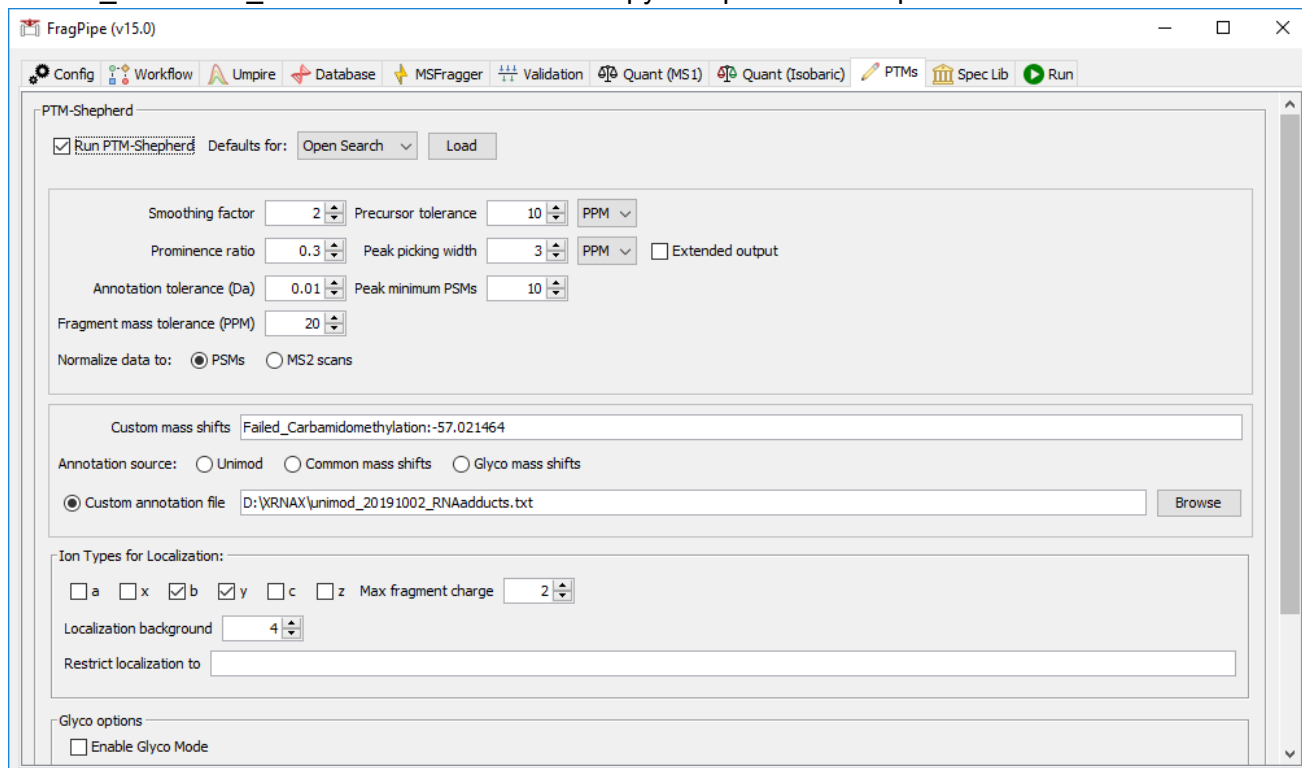
Name	Date modified	Type
bin	3/9/2021 11:08 PM	File folder
cache	3/9/2021 11:35 PM	File folder
jre	3/9/2021 11:08 PM	File folder
lib	3/9/2021 11:08 PM	File folder
tools	3/9/2021 11:22 PM	File folder
updates	3/30/2021 11:01 PM	File folder
workflows	3/9/2021 11:08 PM	File folder

2. Close and then re-launch FragPipe to refresh the workflows. On the Workflow tab, select XRNAX-MassOffset from the workflow dropdown menu and press ‘Load’. Drag and drop or use ‘Add

files' to input the file 20160530_QE1_JT_XRNAXpep_Trp_hydro.mzML.



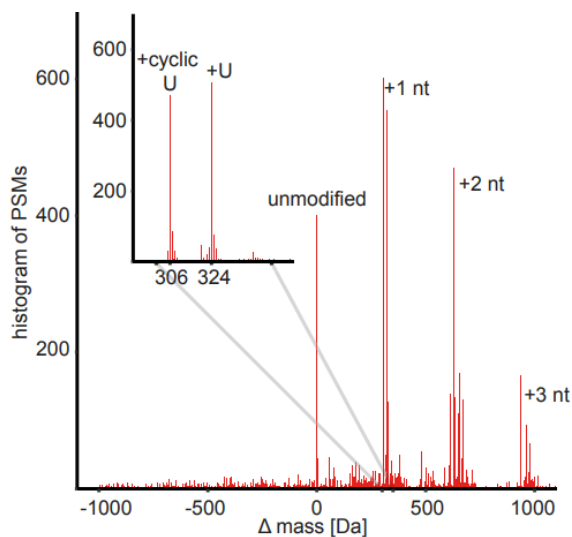
3. On the MSFragger tab, inspect the search parameters. Scroll down to the Mass Offsets box in the Advanced Options section-- this is where the crosslinked RNA fragment masses are specified.
4. Verify that a FASTA sequence file is set on the Database tab. The same human database from the previous analyses can be used (or you can choose to download a new one).
5. On the PTMs tab, we will add a custom annotation file to supply names for the mass offset values (as a list of name-mass pairs). In the 'Custom annotation file' box, provide the file path to unimod_20191002_RNAadducts.txt. You can copy and paste the file path or 'Browse' to add it.



6. Now we can start the analysis from the 'Run' tab after specifying an output location, you can make a new folder: 'fragpipe_RNA-crosslinking_results'

Examine the results

- Inspect the global.profile.tsv file. Explore the results. Do you observe the mass shifts discussed in the manuscript (Figure 2B, shown below)? What are the mapped_mass_1 names given to the top two entries? Do these two mass shifts each predominantly localize to a particular residue(s)? Does the annotation (mapped_mass_1 name) of the top mass shift show up elsewhere in this file?



- Look through the Protein Descriptions of the identifications in the protein.tsv file. The Razor Observed Modifications column lists all the mass shifts from PSMs mapping to each identification.