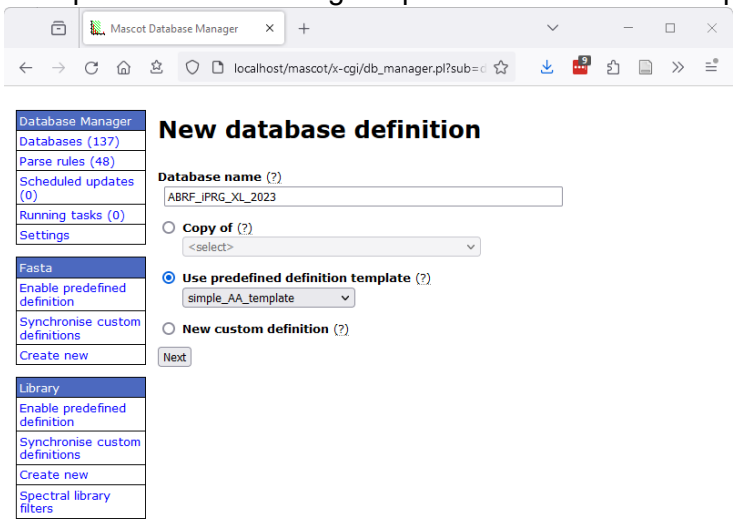
ABRF iPRG 2023 crosslinking study tutorial

The aim of this tutorial is to enable participants in the 2023 ABRF iPRG crosslinking study to analyze the study data set with Mascot Server. This tutorial will walk you through the steps required to: (a) prepare and configure Mascot Server for this data set; (b) process the data automatically with Mascot Server and Distiller; (c) evaluate the crosslinking results; (d) export the results to xiVIEW for further analysis and visualization. There will be a supplementary video that goes into more detail for the different steps. Data processing requires access to Mascot Server, version 2.7 or above, using either a local copy of Mascot Server or the public server (see below). You can also find links to mgf peak lists and search results in the resources section at the end of this document in case you want to skip ahead to the crosslink analysis.

In-house Mascot Server method set up

1. *Create a custom database of the target proteins*. A FASTA file (ABRF_iPRG_XL_2023.fasta) has been provided that contains the proteins of interest. This is a combined database of E. coli proteins, from the expression system, common contaminants and the expressed proteins. Mascot Server reports the best match in the database for each query. Ideally, we want to match MS2 spectra from these background proteins so that they cannot become false positives in the crosslinking analysis.

   Set up the database using the predefined definition template “simple_AA_template”:

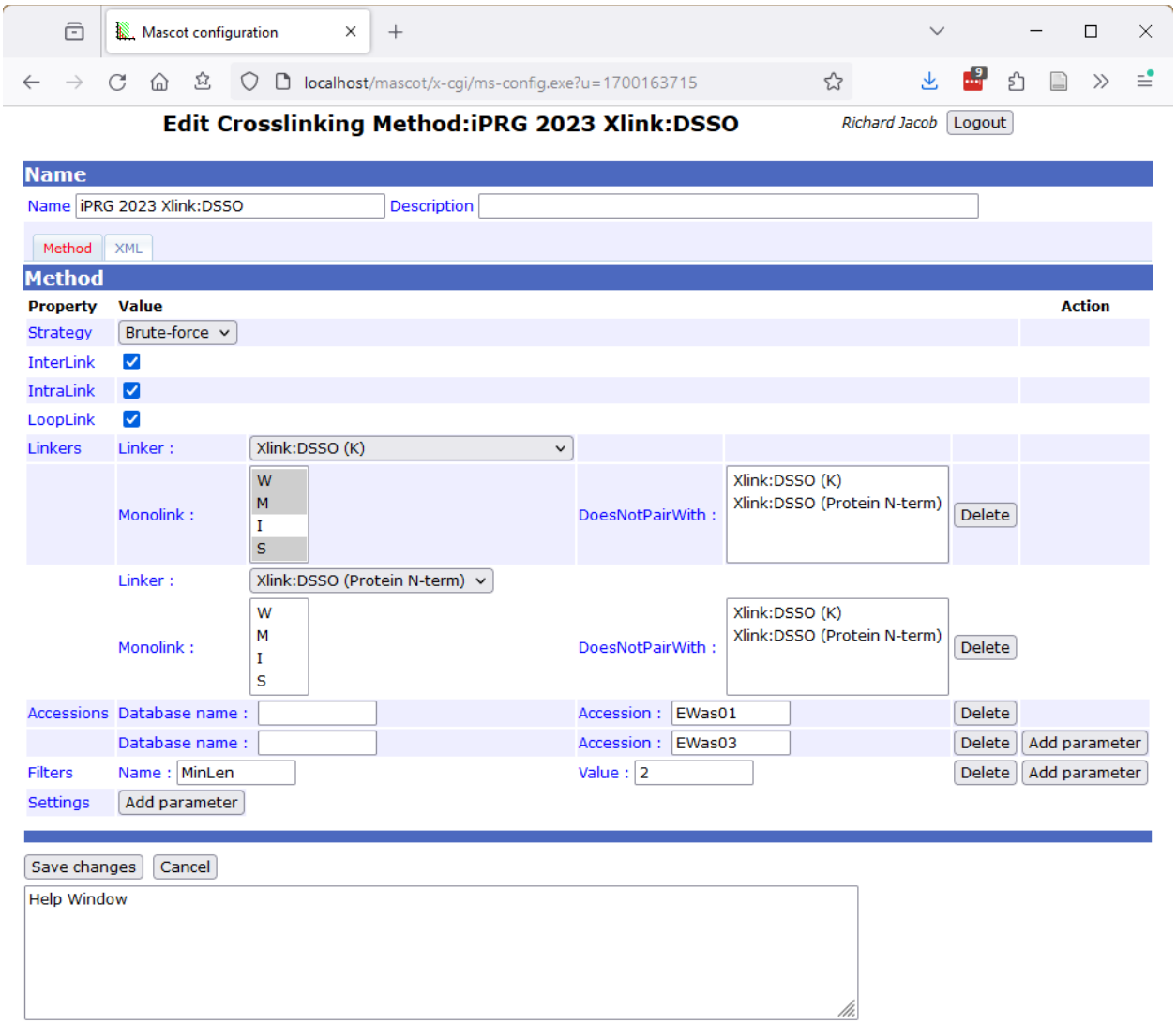


   Step through the dialogs accepting the defaults and chose to “Upload the file using web browser”. Once the fasta file has been uploaded “Activate” it.

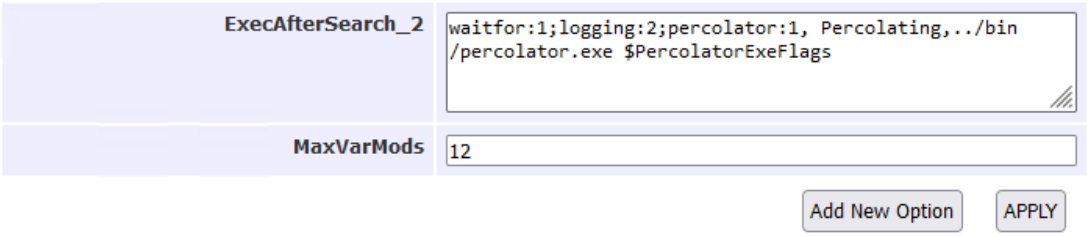
2. *Configure the crosslinking method.* The crosslinking agent used for these analyses, DSSO, is included in the default definitions on Mascot Server and can be viewed in the “Linkers” editor. In the Configuration editor->Crosslinking, click on "New crosslinking method" to add a new crosslinking method.

   Configure the method to use DSSO as the crosslinking agent. For the Monolink fields, select “W, M, S, T, A, R” for the lysine (K) linker. These monolinks correspond to various quenched crosslinking reactions and fragments related to the chemistry of the reagents. Do not select any monolinks for the Protein N-term. At this stage you also need to specify the accession

numbers of the proteins that are being investigated (included in the custom database ABRF_iPRG_XL_2023.fasta). If you don't know the accession numbers, you can search one of the files without a crosslinking method and try to determine which accession numbers to use.



3. *Increase maximum number of variable modifications allowed in a search.* Mascot Server has a default value of 9 variable modifications allowed in a search. In a crosslinking search each of the Monolinks counts towards the number of variable modifications along with the crosslink modifications themselves and any other modifications selected in the search form. In the Configuration editor->Configuration options editor search for the MaxVarMods parameter. If it exists then make sure is set to 12 or more. If it does not exist scroll down to the bottom of the page and add a new option with a value of 12 and apply the changes.



Processing and searching the data

4. *Raw file data processing and search options.* The raw files can be processed as described below: (a) independently and searched through the web interface, (b) processed in Mascot Distiller and then searched or (c) processed automatically with Mascot Daemon. Crosslinked peptides can have quite high masses which means they often have higher charge states too. When picking the peaks either decharging the fragment ions to MH+ or reporting their charge states helps a lot in the database search. This is discussed in depth in a previous article. There are three possible ways to do this.

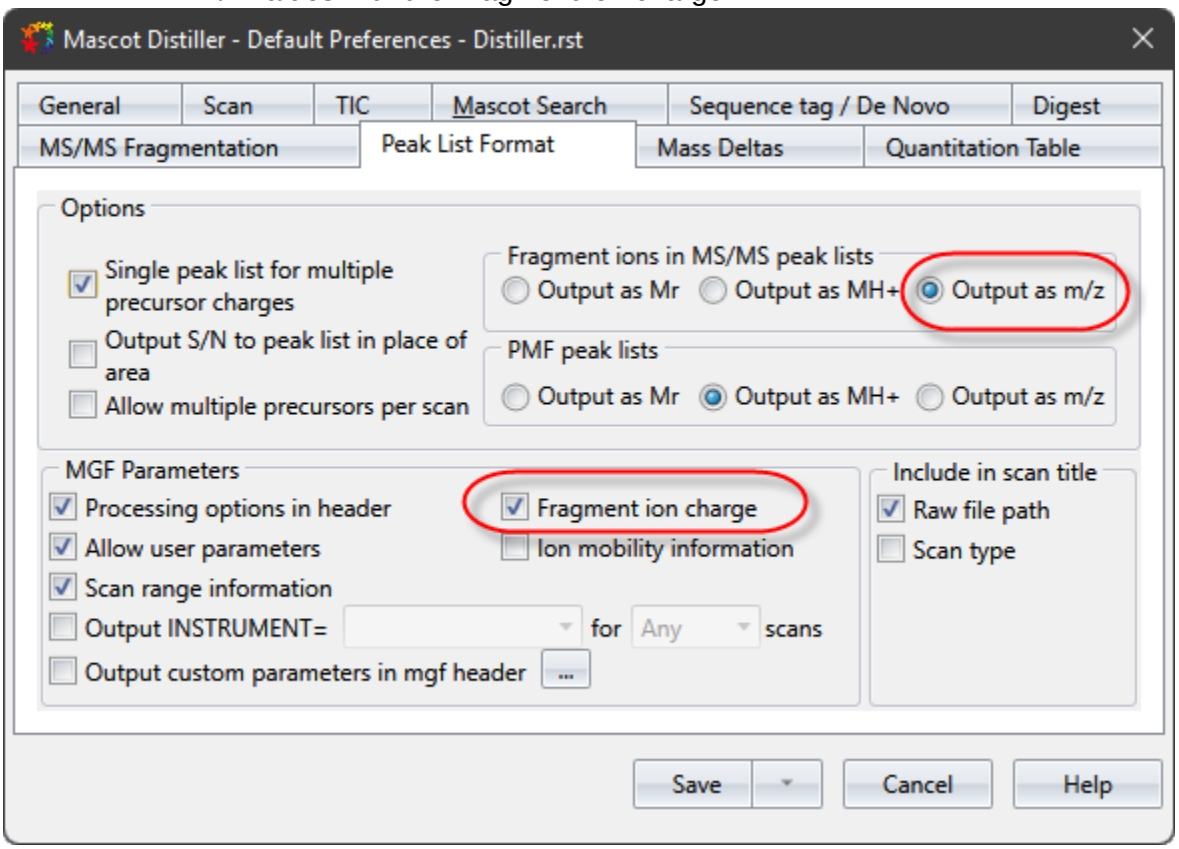
    a. *Searching a peak list directly*. Use your favorite peak picking software to generate a peak list from the raw data in a format suitable for Mascot Server. The MGF and mzML formats are the most common ones. We also have pre made MGF files available. Open up a MS/MS ion search form on a Mascot Server, here is a link to the public server, and set the search conditions with suitable databases, crosslinking method and measurement tolerances. Here are the ones I used:

---Or---

b. *Using Mascot Distiller.* Mascot Distiller can be used for peak picking and submitting a search. If you would like to try Mascot Distiller, a 30-day trial is available. If you process the data using the prof_prof.ThermoXcalibur.opt options peak picking will take about three hours per file. I modified the settings to use 200 data points per Dalton reducing processing time to about half an hour per

file. You can download these modified settings here. You can also use Mascot Distiller with the default.ThermoXcalibur.opt options and processing time will be less than 10 minutes or so per a file but there will be no charge state information for the fragment ions which can be an issue for larger crosslinked peptides.

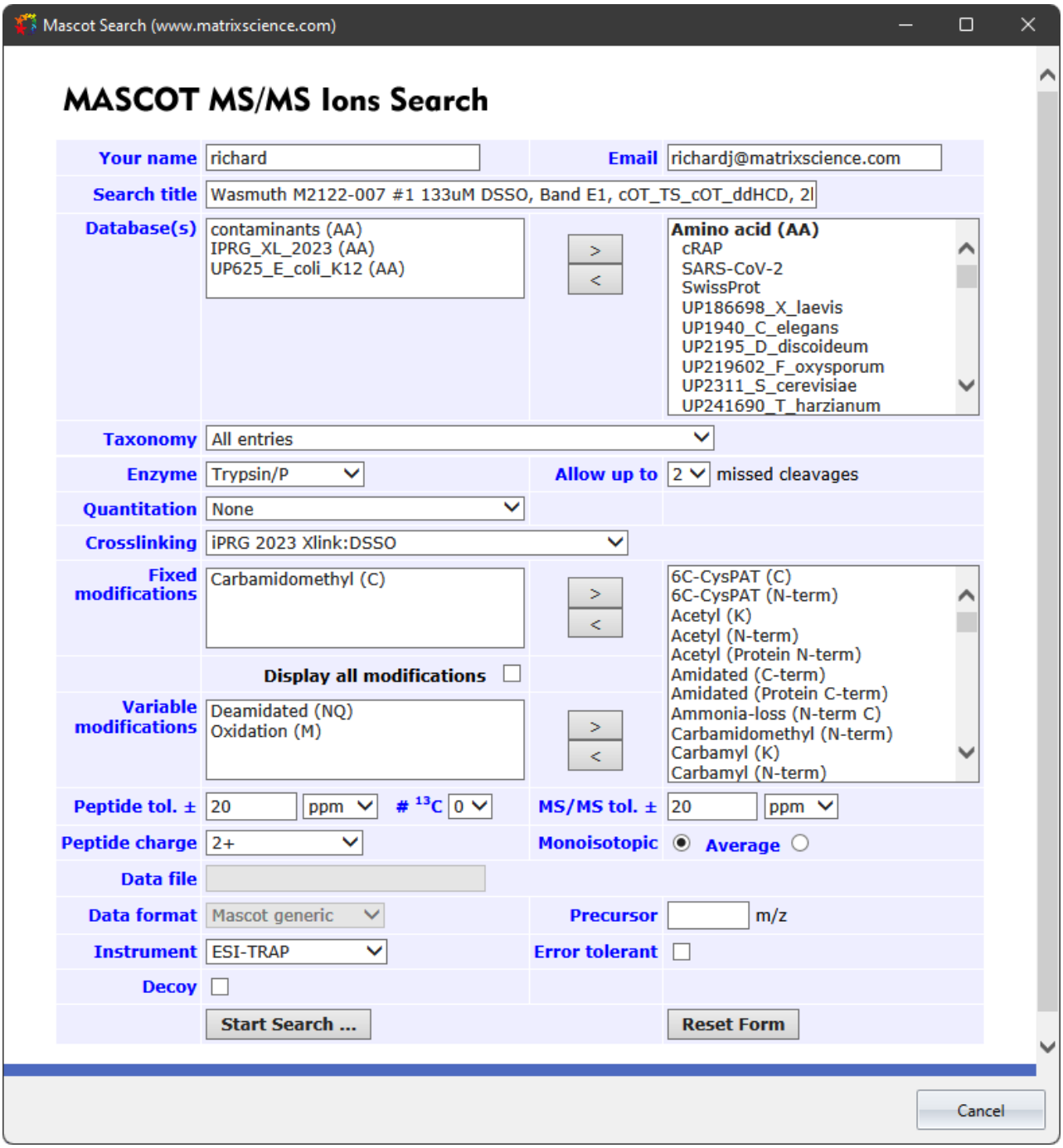
i. Before processing the raw data change the Mascot Distiller->Tools->Preferences->Peak List Format settings to save the fragment ions as m/z values with the Fragment ion charge:



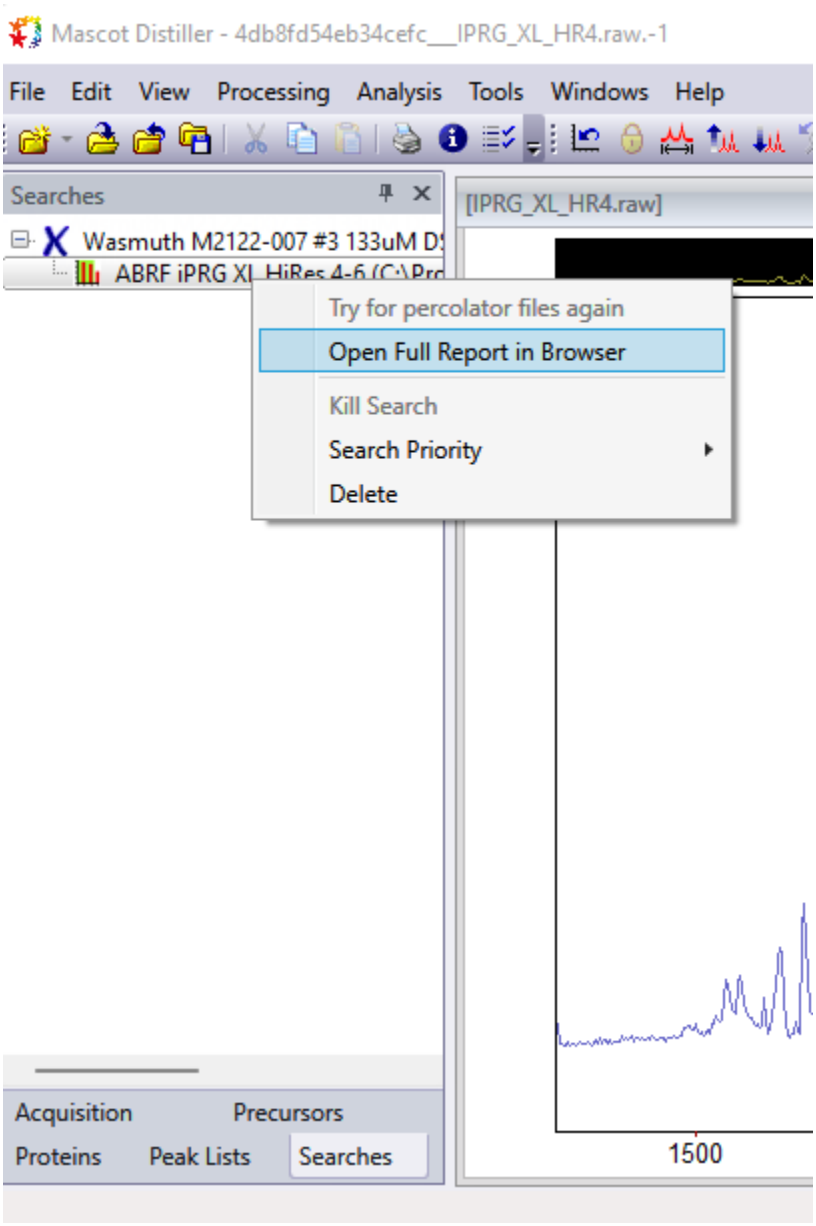
ii. In Mascot Distiller open a data set from the File->New Project->Thermo dialog.

iii. Set the Mascot Distiller processing options from the Processing->Edit Processing options->Load dialog box.

iv. Start the data processing from the Processing->Process And Search menu. Set up the search parameters and Start Search.

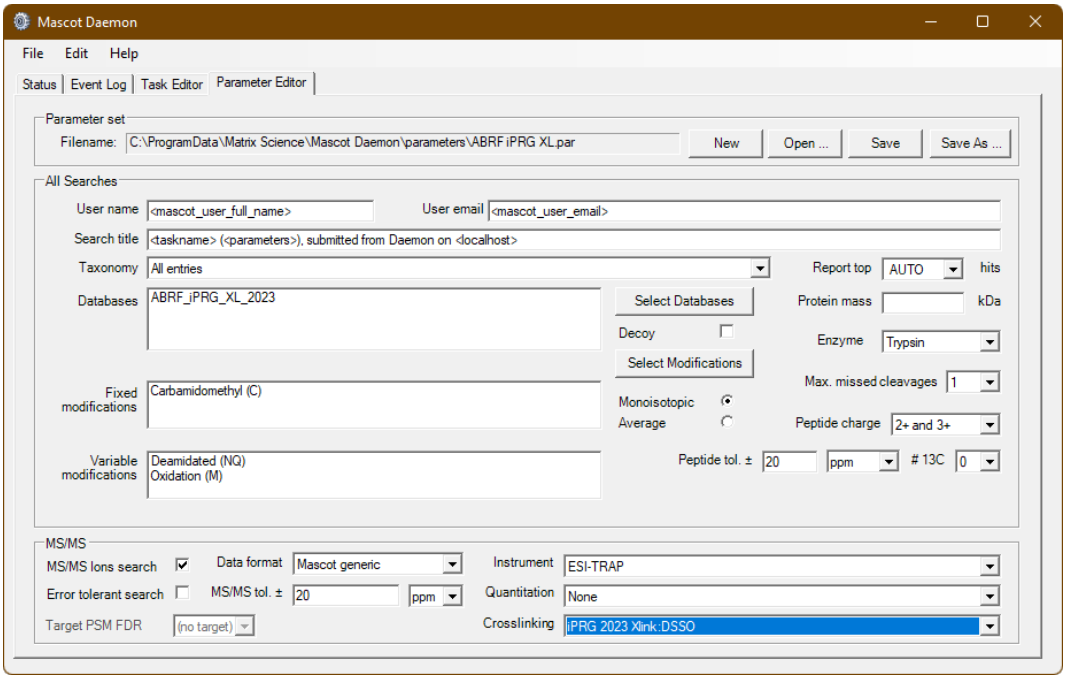


v. Save the search project once the search is complete.
vi. Open the search results in a web browser.

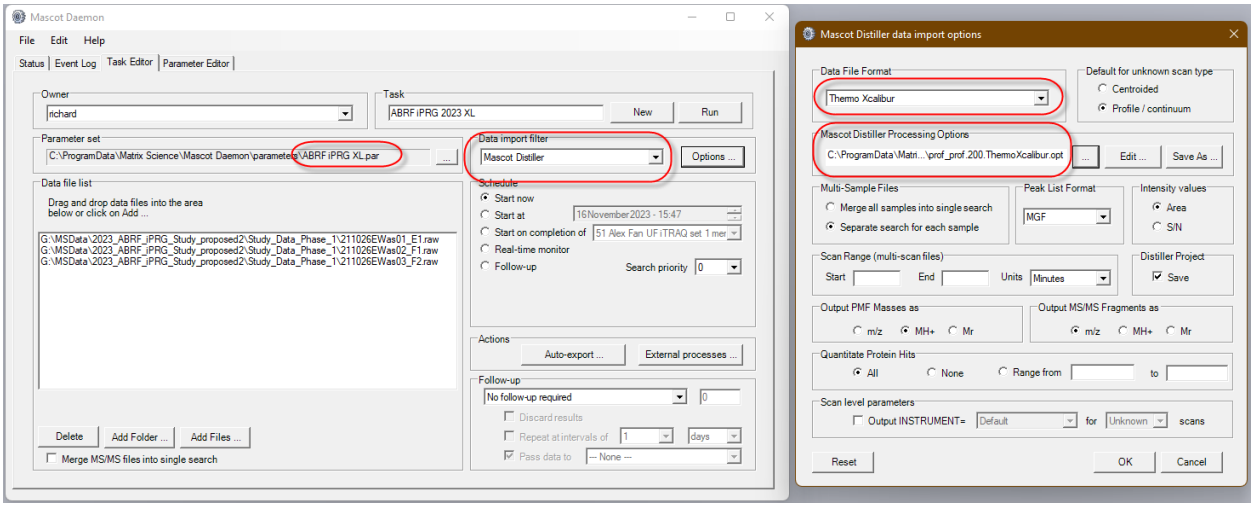


---Or---

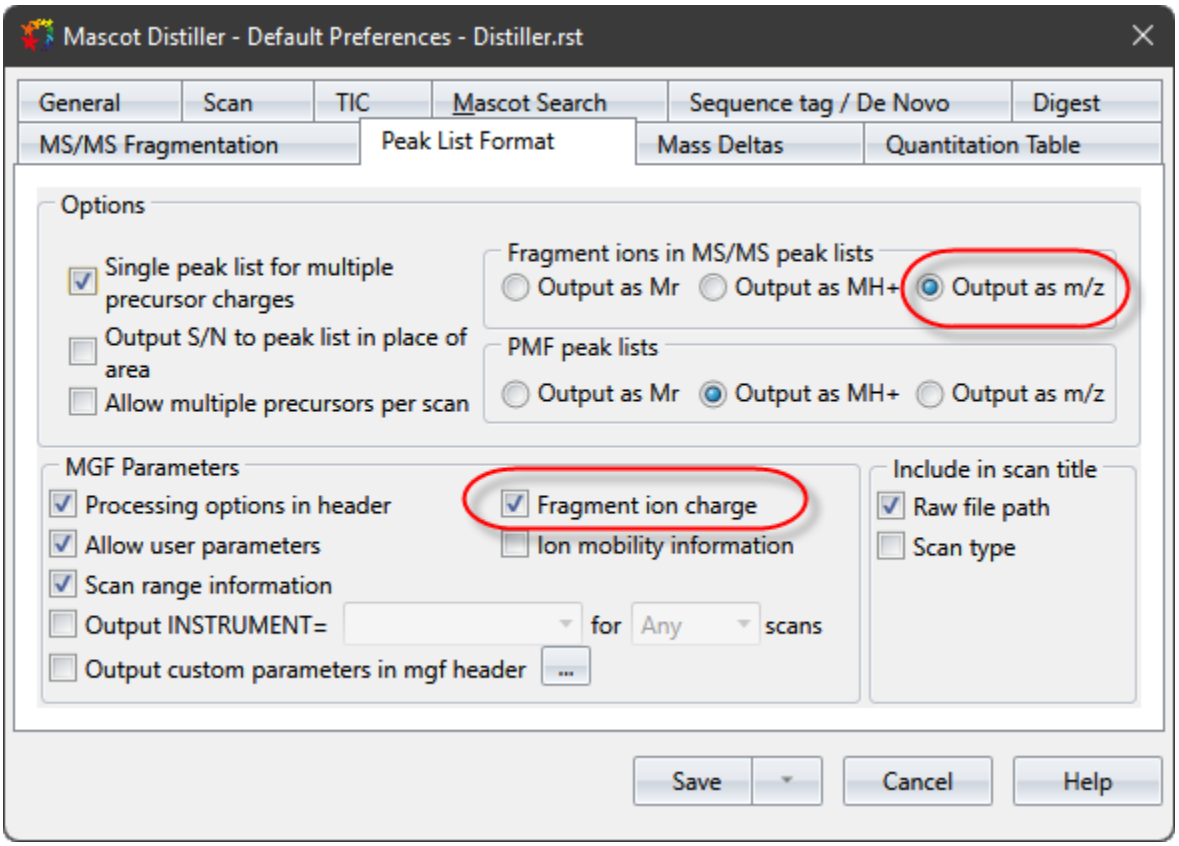
c. *Automated processing with Mascot Daemon*. You'll need a local Mascot Server installation that includes Mascot Daemon for this option. Here I am using Mascot Distiller for the data import filter but you can also search peak lists that have been previously generated or use one of the other data import filters. Quick note: If you have just added the crosslinking method to Mascot Server you will need to restart Mascot Daemon in order to see the new method.
    i. Set up the search parameters file, these are the parameters I used:



ii. As with using Mascot Distiller directly I modified the peak picking parameters to use less data points per Dalton reducing processing time to about half an hour per file. You can download these modified settings here.



iii. If the data is profile data either by acquisition or by uncentroiding and you are using a prof_prof.ThermoXcalibur.opt based processing options we need to make sure we can use the charge state information from the fragment ions. Either set up the Output MS/MS fragments as MH+ in the data import filter or configure Mascot Distiller to include the fragment ion charge state in the peak list. Open Mascot Distiller without a data file and go to Mascot Distiller->Tools->Preferences->Peak List Format settings to save the fragment ions as m/z values with the Fragment ion charge. Close Distiller once edited. If you are using the default.ThermoXcalibur.opt options or other options that do not process the data in profile mode then the fragment ion charge will not be determined and you don't need to modify the peak list format settings.



iv. Run the task to process and search the data.
v. Open the search results in a web browser.

Evaluating the search results

5. At this point you should be viewing the search results for one of the files in a web browser. Scroll down a little and click on the black arrow to expand “Modification statistics for all protein families” to see a quick overview of the number of matches in the data:

Modification statistics for all protein families

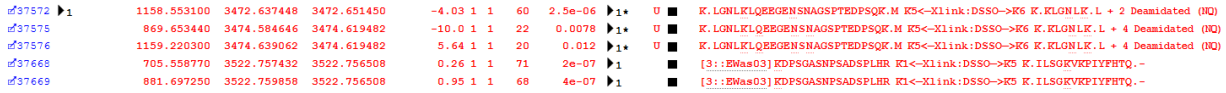
Modification	Delta	Type	Site	Total matches
Carbamidomethyl	57.021464	fixed	C	3143
Oxidation	15.994915	variable	M	2707
Deamidated	0.984016	variable	N	739
Deamidated	0.984016	variable	Q	412
Xlink:DSSO[W]	176.01433	variable	K	243
Xlink:DSSO	158.003765	crosslink	K - K	235
Xlink:DSSO[R]	279.077658	variable	K	126
Xlink:DSSO[A]	54.010565	variable	K	35
Xlink:DSSO[M]	175.030314	variable	K	34
Xlink:DSSO[S]	103.9932	variable	K	14
Xlink:DSSO[T]	85.982636	variable	K	11
Xlink:DSSO	158.003765	crosslink	K - Protein N-term	7

This table is taken from 211026EWas01_E1.raw report. There are 235 K to K crosslinks and 7 K to protein N-terminal crosslinks reported in the search along with many oxidized methionines, deamidations plus monolinks from the crosslinking reagent.

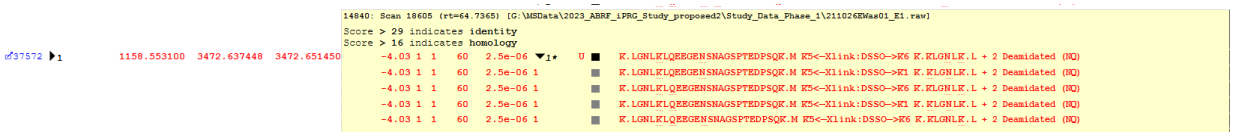
Scroll down to the proteins section of the report and click on a black triangle next to the

protein name for the first hit for one of the target proteins. EWas01 in this report:

This will expand the list of peptide matches for that protein family. Crosslinked peptides have two peptide sequences with the crosslinking agent, <—Xlink:DSSO—>, between them. Matches with high scores/lower expect values are the most reliable. The report below gives an indication of any match level ambiguity with an asterisk by the top 10 matches expander in the middle of the line.



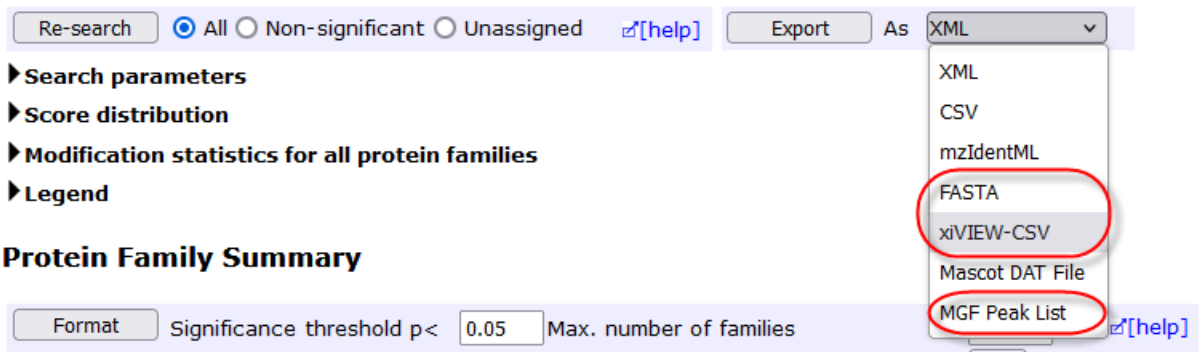
This means that there is another arrangement or match that has the same or very similar score:



The crosslink matches can be viewed by clicking on the query number to see the peptide view and an annotated spectrum. We have previously published some guidelines for validating intact crosslinked peptide matches and you can use those guidelines to determine what a good match looks like and which ones don't have sufficient information to validate the crosslink or are poor/random matches. With a large number of matches it can be difficult to see the relationship between the crosslinked matches due to multiple charge states, modifications and different peptide lengths. We find it easier to view the results in specialized software designed for the task (see below).

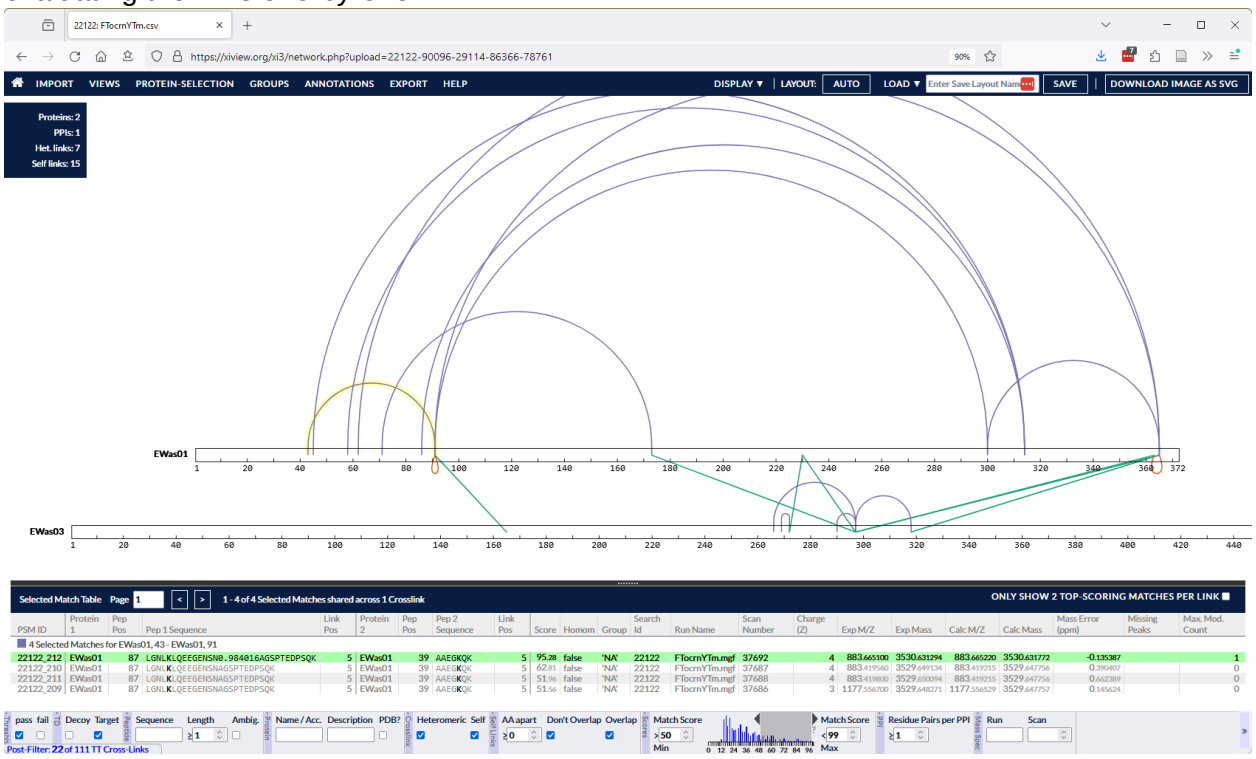
Exporting to xiVIEW for crosslink validation and visualization

6. You can export the results for xiVIEW using the xiVIEW CSV, the mgf peak list and the protein sequences in FASTA format. It is important to export the mgf file with the queries in "Mascot query order". The results are exported by sequentially specifying: "xiVIEW-CSV," "FASTA" and "MGF Peak List" (i.e., as separate exports):



Select the export format and click the export button and follow the dialog boxes to complete the export. Do the same thing for the other two reports. Full details on how to prepare the data for xiVIEW are here.

7. The exported results can then be imported into xiVIEW and visualized showing the intra and inter bonds between the different proteins. You can use the validation guidelines to review the individual links in xiVIEW either filtering the results based on a minim score cut-off or evaluating the links one-by-one.



Looking at the results I did not see many crosslinked peptides that had sufficient evidence to support the identification with a score below 50. Here I have applied a minimum score of 50 as an arbitrary cutoff for a crosslinked peptide matches. A better approach to filtering on the score would be to filter on the expect value but XiVIEW does not support that yet.

What if you don't have a local copy of Mascot Server?

As part of the ABRF iPRG study, we have added the iPRG protein sequences and crosslinking method to the public server. However, the public server has limits to the number of queries it can analyze in a single search. The data files for the iPRG study are too large for a public search. If you would like to try searching this dataset on the public Mascot Server, please contact support@matrixscience.com to request a temporary account with expanded limits.

The goal of this tutorial is to show you how easy it is to search and analyze a crosslinked protein data with an in-house installation of Mascot Server or on the public server. The actual searches are very fast but it can take a lot of time to fully analyze the results. The use of xiVIEW does facilitate the process.

We have also created a walkthrough video of the analysis covering the points mentioned above but going into more detail for the individual steps. As always if you have any questions please contact support@matrixscience.com

Resources

Filename	MGF file	Mascot Server search result	xiView project
211026EWas01_E1.raw	211026EWas01_E1.mgf	FTocrnYTm.dat	FTocrnYTm.csv
211026EWas02_F1.raw	211026EWas02_F1.mgf	FTocrnYOO.dat	FTocrnYOO.csv
211026EWas03_F2.raw	211026EWas03_F2.mgf	FTocrnYOS.dat	FTocrnYOS.csv

Mascot Distiller raw data processing options prof_prof.200.ThermoXcalibur.opt
*The raw data can be obtained from the study invitation.