

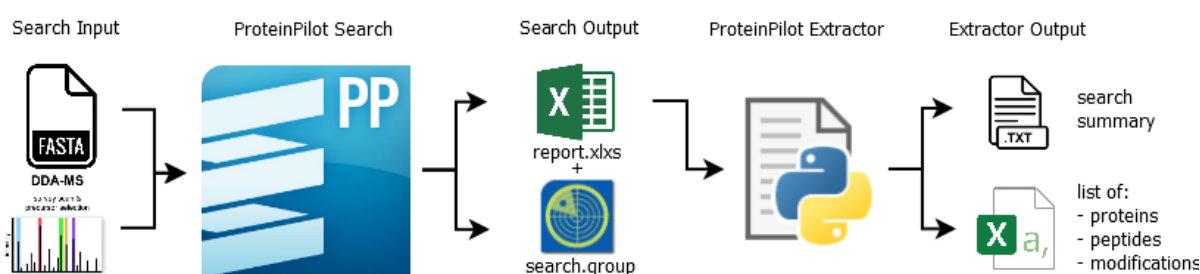
## ProteinPilot Report Data Extractor

[Sciex ProteinPilot](#) software is used to search DDA proteomics data against a FASTA database file. Search results can be used for individual analysis or generation of SRLs used for processing DIA/SWATH data.

ProteinPilot searches generate a report file in excel format that contains key search results. Of the two types of reports that can be generated (full or brief), the full report can take anywhere between 5-10 minutes to load in excel. This step is not limited by computational power alone but more by excel itself.

This report extractor pulls the most frequently queried information from the search results out of the excel file and saves it in a text file. Although this extraction also takes up to 5 minutes, multiple result reports can be extracted sequentially and the text outputs loads instantaneously, saving significant time if result files need to be re-opened.

The extractors integration into a data analysis pipeline is summarised in the workflow below:



The extractor can be run in Spyder or as a standalone .bat batch script, if python is installed. The .bat will extract all the report files in the folder where it is placed. The Spyder script will do the same, however, is useful if you wish to only extract one file in a specified directory. The Spyder script contains instruction on how to run.

The following information is included in the search summary txt output:

- Number of spectra in search
- Protein 1% Global FDR
- Peptide 1% Global FDR
- Peptide Spectrum Match 1% Global FDR
- Digestion Efficiency (percentage displayed as decimal) for:
  - o Canonical sequence peptides
  - o Over-cleaved sequence peptides
  - o Under-cleaved sequence peptides
  - o Mixed over and under cleaved
- (Top 15) Most Frequent Modifications
- Correspondence between FDR Levels and ProteinPilot Reported Confidences (as % Global FDR from Fit)
- Chromatographic Peak Width at Half Height (Median in sec)
- Database Filename used
- Cysteine Alkylation setting
- Species

In addition to the search summary, a .csv file is generated containing the list of peptides and proteins identified at 1% FDR, as well as the complete list of peptide modifications and their identified frequency. Note that the way the output is formatted, there is no correlation between the peptides and proteins. This output is useful for further analysis, such as calculating the physical properties of the peptides identified, or protein list comparison and gene ontology analysis.

Examples of these inputs and outputs can be found in the GitHub repository [@stevew91](#).