MS Ana Spectral Library Search for Proteome Discoverer 2.1

User Manual

July 2018

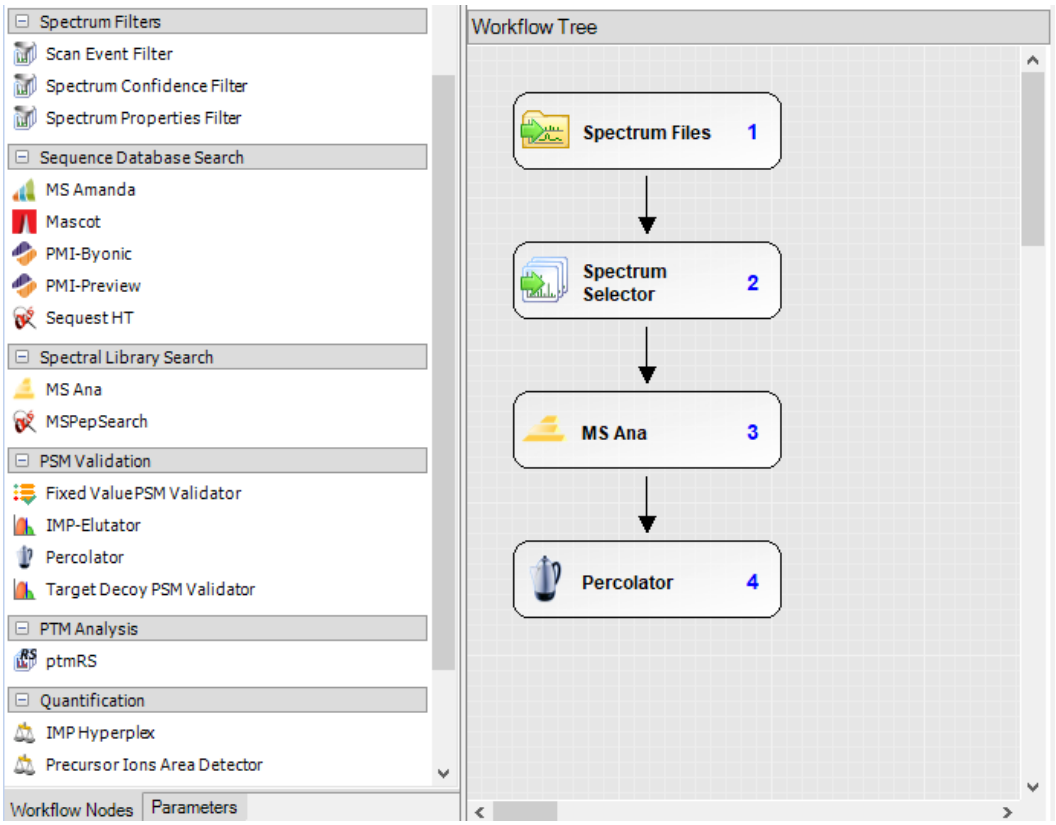
MS Ana is a scoring system to identify peptides out of tandem mass spectrometry data using libraries of previously identified spectra. Furthermore, MS Ana can create decoy spectral libraries to allow for the validation of results using the target-decoy approach.

Installing MS Ana

Make sure Proteome Discoverer 2.1 is closed. Then, run the MS Ana setup file **MSAnaPD21.exe** and accept the license agreement. The setup will copy all relevant files into the PD 2.1 directory and register the node with Proteome Discoverer. Wait until the console window closes then hit Finish to exit the setup. The MS Ana Spectral Library Search Node is now ready to use! ***Please note: If you are experiencing problems during the installation because of Windows security services you need to use the provided certificate to add FHOÖ & IMP as a trusted software source. Double-click the*** **FHOOE_IMP_CERT.cer*****, then choose 'Install Certificate...' and follow the instructions on screen.***

Running MS Ana

Using MS Ana in Proteome Discoverer is similar to using Mascot or SEQUEST. Simply drag & drop the MS Ana node into your workflow and connect it to its predecessor and successor nodes.

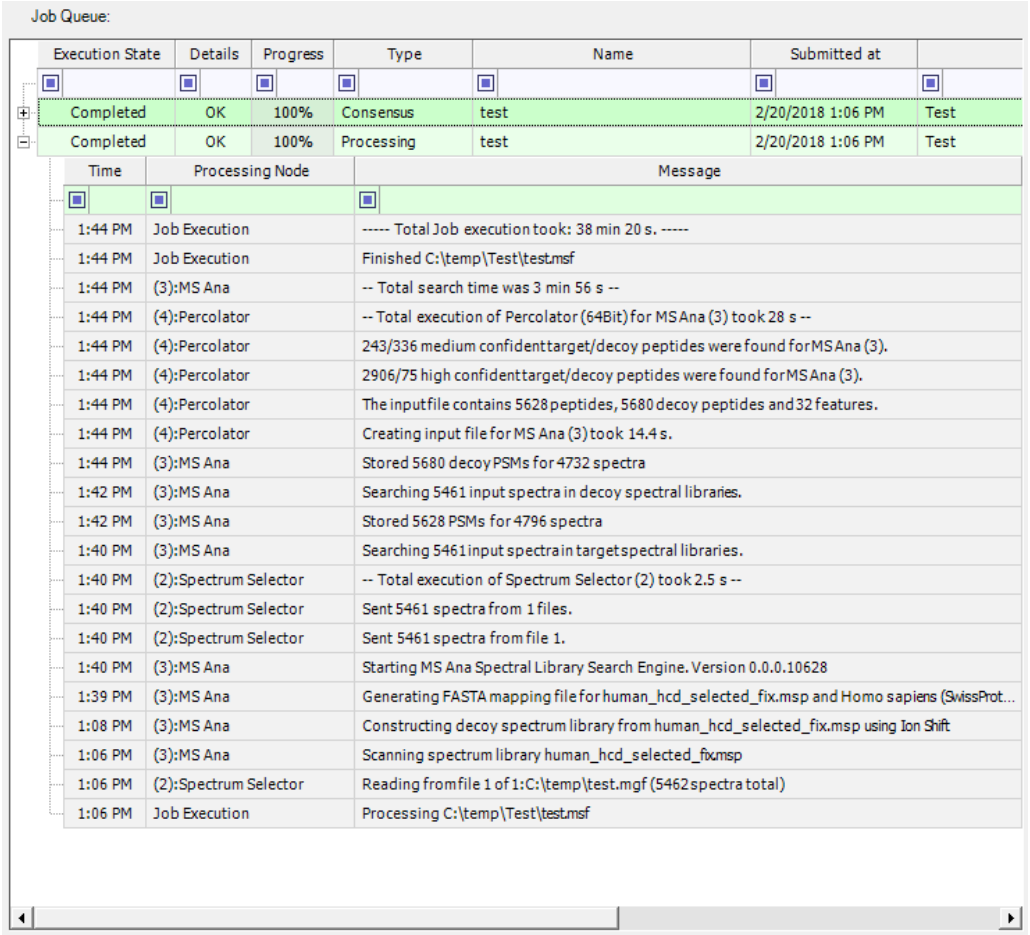


At a minimum, you will need to select a spectral library in MSP format that was previously registered to Proteome Discoverer. Afterwards, please set the parameters (described in the following section) according to your experiment composition and click the run button.

Alternatively, you may use the provided example analysis template. In the Study View select 'Open Analysis Template' and navigate to the .pdAnalysis file provided with the MS Ana installation. This will load appropriate processing and consensus workflows.

Please note: The first time you run a search with a spectral library, MS Ana will create the decoy library to allow result validation using the target-decoy approach. This will take considerably more time than the library search (depending on library size). Subsequent runs will be much faster since the decoy library is stored and used again.

After the workflow is finished, you can view the result report.



Job Queue:

Execution State	Details	Progress	Type	Name	Submitted at	
Completed	OK	100%	Consensus	test	2/20/2018 1:06 PM	Test
Completed	OK	100%	Processing	test	2/20/2018 1:06 PM	Test

Time	Processing Node	Message
1:44 PM	Job Execution	----- Total Job execution took: 38 min 20 s. -----
1:44 PM	Job Execution	Finished C:\temp\Test\test.msf
1:44 PM	(3):MS Ana	-- Total search time was 3 min 56 s --
1:44 PM	(4):Percolator	-- Total execution of Percolator (64Bit) for MS Ana (3) took 28 s --
1:44 PM	(4):Percolator	243/336 medium confident target/decoy peptides were found for MS Ana (3).
1:44 PM	(4):Percolator	2906/75 high confident target/decoy peptides were found for MS Ana (3).
1:44 PM	(4):Percolator	The input file contains 5628 peptides, 5680 decoy peptides and 32 features.
1:44 PM	(4):Percolator	Creating input file for MS Ana (3) took 14.4 s.
1:44 PM	(3):MS Ana	Stored 5680 decoy PSMs for 4732 spectra
1:42 PM	(3):MS Ana	Searching 5461 input spectra in decoy spectral libraries.
1:42 PM	(3):MS Ana	Stored 5628 PSMs for 4796 spectra
1:40 PM	(3):MS Ana	Searching 5461 input spectra in target spectral libraries.
1:40 PM	(2):Spectrum Selector	-- Total execution of Spectrum Selector (2) took 2.5 s --
1:40 PM	(2):Spectrum Selector	Sent 5461 spectra from 1 files.
1:40 PM	(2):Spectrum Selector	Sent 5461 spectra from file 1.
1:40 PM	(3):MS Ana	Starting MS Ana Spectral Library Search Engine. Version 0.0.0.10628
1:39 PM	(3):MS Ana	Generating FASTA mapping file for human_hcd_selected_fix.msp and Homo sapiens (SwissProt...
1:08 PM	(3):MS Ana	Constructing decoy spectrum library from human_hcd_selected_fix.msp using Ion Shift
1:06 PM	(3):MS Ana	Scanning spectrum library human_hcd_selected_fix.msp
1:06 PM	(2):Spectrum Selector	Reading from file 1 of 1:C:\temp\test.mgf (5462 spectra total)
1:06 PM	Job Execution	Processing C:\temp\Test\test.msf

Parameters of MS Ana

The following parameters can be set for configuring the MS Ana scoring system:

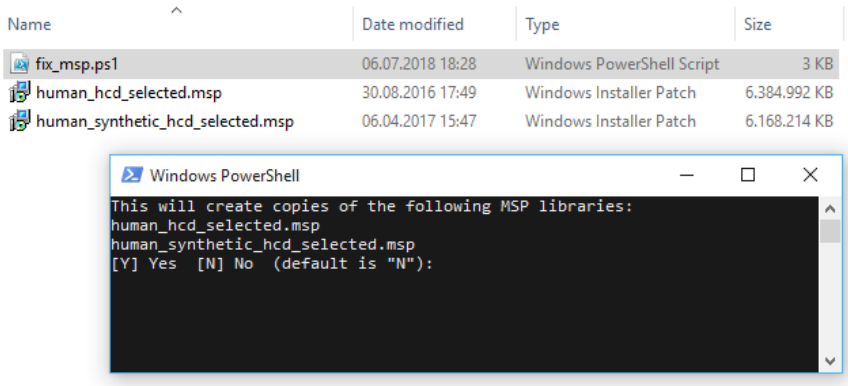
1. **Input Data**
    a. **Spectral Library**
    The spectral library or libraries that will be used by MS Ana for peptide identification. Please select any library or libraries (in MSP format) previously added using the Proteome Discoverer Spectral Library interface.
    Please note: The first time you run a search with a spectral library MS Ana will create the decoy library to allow result validation using the target-decoy approach. This will take considerably more time than the library search (depending on library size). Subsequent runs will be much faster since the decoy library is stored and used again.
    b. **Protein Database**
    The protein sequence database(s) used to map the identified peptides to proteins. Please select any FASTA database(s) previously added using the Proteome Discoverer FASTA File Interface.
    Whenever an identified peptide is not part of any selected protein database (or no protein database is selected) MS Ana will report the protein accessions deposited in the spectral library file (if any are available).
    c. **Decoy Method**
    Specify which method should be used for the creation of decoy spectral libraries:
        i. **Ion Shift *(Recommended)***
        Constructs decoy spectra by shifting peptide sequences and moving annotated fragment ion peaks accordingly. Annotation of fragment ion peaks will use the selected Fragment Tolerance.
        ii. **Precursor Swap**
        Constructs decoy spectra by swapping precursor information leaving fragment ion information intact. Creation is potentially faster but can cause issues for small libraries in particular leading to underrepresented decoy hits.
2. **Search Settings**
    a. **Monoisotopic Precursor**
    Define whether monoisotopic mass should be used instead of average mass for the precursor ion.
    b. **Precursor Tolerance**
    The tolerated mass error when matching precursor mass values (MS1 mass error). Possible units are “Da” and “ppm”.
    c. **Fragment Tolerance**
    The tolerated mass error when matching fragment ion mass values (MS2 mass error). Possible units are “Da” and “ppm”.
    d. **Shown Ranks**
    Specify how many hits per spectrum should be reported (beyond the best one).
3. **Preprocessing**
    a. **Query Peak Picking**
    Specify the depth for filtering low intensity peaks in the query spectra during preprocessing. Only the N highest peaks per 100 m/z window are kept.

b. **Library Peak Picking**
   Specify the depth for filtering low intensity peaks in the library spectra during preprocessing. Only the N highest peaks per 100 m/z window are kept.
c. **Remove Precursor**
   Define whether to remove all fragment ion peaks corresponding to the precursor or its neutral losses during preprocessing.
d. **Minimum Peaks**
   Any query spectra with total peak count less than the specified value are ignored during the spectral library search.

Adding Spectral Libraries to Proteome Discoverer 2.1

To search in a spectral library using MS Ana it must be added using the Spectral Library interface. **Currently, MS Ana only supports libraries in MSP Format.** Suitable libraries for different standard organisms can be downloaded from the NIST website at http://chemdata.nist.gov/. In Proteome Discoverer 2.1 go to Administration → Maintain Spectral Libraries → (+) Add and select the .MSP file to start importing the spectral library. For large libraries, this will take a considerable amount of time while the library is copied and registered to PD services.

Please Note: The latest spectral libraries from the NIST website cannot be imported in Proteome Discoverer 2.1 because of incompatible formatting. We provide a quick fix for this issue in the form of a Windows PowerShell Script. Copy the* fix_msp.ps1 *from the MS Ana installation folder into the directory with the spectral libraries. Right-click the* fix_msp.ps1 *and select "*Run with PowerShell*" to execute the script. A dialog will ask for confirmation before the script creates copies of the libraries with corrected formatting. Progress is displayed in the PowerShell window and a notification appears once the process is finished. You can then add the newly created *_fix.msp *library file to Proteome Discoverer.



Contact

This research project is a collaboration of the Protein Chemistry Group at IMP and the Bioinformatics Research Group at FH Upper Austria, Hagenberg Campus. For any further questions, bug reports, or ideas please contact Sebastian Dorl, Viktoria Dorfer, Stephan Winkler, Karl Mechtler, or post your comment in the IMP Nodes for Proteome Discoverer Google Group.