glyXtool^{MS} Usermanual

version 2.0

2018-08-11

Contents

1 Installation

1.1 OpenMS

For installation of OpenMS visit https://www.openms.de/ and follow the download/install or build instructions for your operating system.

After installation the following tools should be installed: TOPPAS and TOPPView.

1.2 Python

To run glyXtool^{MS}, a python 2.7 installation is required, together with the package manager pip. The use of a virtual environment like virtualenvwrapper is recommended if other python versions are/will be installed on the same workstation.

Install python 2.7 from https://www.python.org/. The package manager for python will then be installed, too. To check, open a console and type the command “pip”. If it has not been installed, follow the installation instructions on https://pip.pypa.io/en/stable/installing/#do-i-need-to-install-pip.

The use of a virtual environment is recommended, in case multiple python installations with different package setups are installed on the computer. For the installation of virtualenvwrapper, please refer to https://virtualenvwrapper.readthedocs.io/en/latest/

Virtualenvwrapper can be installed via:

```
pip install virtualenvwrapper
```

afterwards a fresh environment can be created using:

```
mkvirtualenv <envname>
```

switch into the environment using:

```
workon <envname>
```

1.3 glyXtool^{MS}

glyXtoolMS can be installed using pip in the command line:

```
pip install glyXtoolMS
```

The dependencies canvasvg, configparser, lxml ,numpy,pyopenms, pyperclip, and xlwt should then be automatically downloaded and installed.

alternatively the .egg or .wheel can be downloaded from https://pypi.org/project/glyxtoolms/

or build manually from https://github.com/glyXera/glyXtoolMS

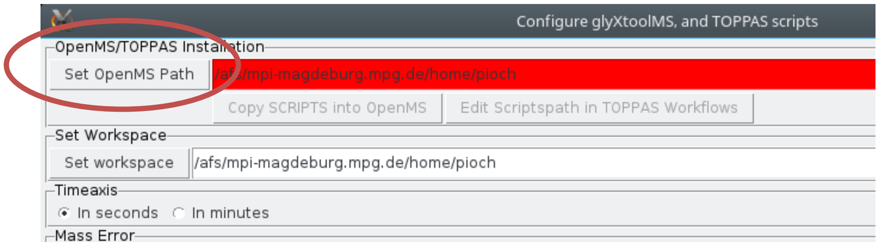
After the installation of glyXtool^{MS}, the glyXtool^{MS} Evaluator should accessible via the console command:

```
glyXtoolMS
```

1.4 TOPPAS Script Setup using glyXtool^{MS} Evaluator

During the first startup of the glyXtoolMS Evaluator, a configuration window for OpenMS will appear(also later available on the Menu/TOPPAS/Configure TOPPAS), since all necessary scripts and tool description files need the be copied from the python package into OpenMS. Please provide the installation path of OpenMS. During the save process all necessary files will be copied over into the OpenMS/share/OpenMS/SCRIPTS/ and OpenMS/share/OpenMS/TOOLS/EXTERNAL/ directories.

Within the same window downloaded TOPPAS workflows (e.g. from the example data sets) can be adapted to the right SCRIPT path.


Configure glyXtoolMS, and TOPPAS scripts
OpenMS/TOPPAS Installation
Set OpenMS Path
Copy SCRIPTS into OpenMS
Edit Scriptspath in TOPPAS Workflows
Set Workspace
Set workspace
Timeaxis
Mass Error

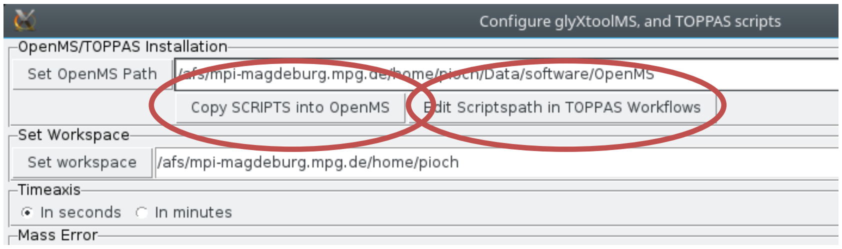
Set OpenMS Path: /afs/mpi-magdeburg.mpg.de/home/pioch

Set workspace: /afs/mpi-magdeburg.mpg.de/home/pioch

Timeaxis: ● In seconds ○ In minutes

Figure 1: OpenMS Path Configuration

The Configuration Panel can be reached duringthe first startup of the glyXtool^{MS} Evaluator, or opened via Menu/Program/Configure. Set the Path to your OpenMS Installation.


Configure glyXtoolMS, and TOPPAS scripts
OpenMS/TOPPAS Installation
Set OpenMS Path
Copy SCRIPTS into OpenMS
Edit Scriptspath in TOPPAS Workflows
Set Workspace
Set workspace
Timeaxis
Mass Error

Set OpenMS Path: /afs/mpi-magdeburg.mpg.de/home/pioch/Data/software/OpenMS

Set workspace: /afs/mpi-magdeburg.mpg.de/home/pioch

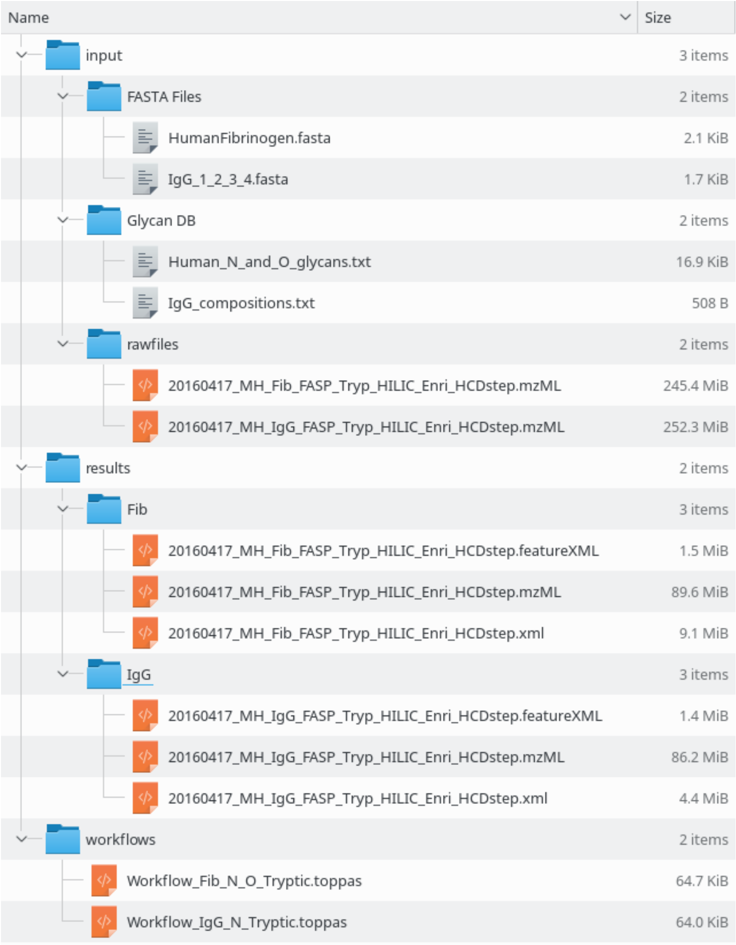
Timeaxis: ● In seconds ○ In minutes

Figure 2: Copy Scripts into OpenMS and edit TOPPAS Workflows

Afterwards the Scripts can be copied over into OpenMS in order to make them available within the TOPPAS pipeline engine. Downloaded TOPPAS workflows can be edited to include the correct scriptpath.

2 Analyzing the Example Data Set

The example data sets of human IgG and human fibrinogen used within the manuscript can be downloaded from https://www.ebi.ac.uk/pride/archive/projects/PXD009716, containing the raw data files, and – within the zip file – the converted raw mass spectrometry files, the FASTA files, the *N*- and *O*-glycan databases, the TOPPAS workflows, the generated reduced mass spectrometry files, the detected features and the analysis files.



Name	Size
input	3 items
FASTA Files	2 items
HumanFibrinogen.fasta	2.1 KiB
IgG_1_2_3_4.fasta	1.7 KiB
Glycan DB	2 items
Human_N_and_O_glycans.txt	16.9 KiB
IgG_compositions.txt	508 B
rawfiles	2 items
20160417_MH_Fib_FASP_Tryp_HILIC_Enri_HCDstep.mzML	245.4 MiB
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML	252.3 MiB
results	2 items
Fib	3 items
20160417_MH_Fib_FASP_Tryp_HILIC_Enri_HCDstep.featureXML	1.5 MiB
20160417_MH_Fib_FASP_Tryp_HILIC_Enri_HCDstep.mzML	89.6 MiB
20160417_MH_Fib_FASP_Tryp_HILIC_Enri_HCDstep.xml	9.1 MiB
IgG	3 items
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.featureXML	1.4 MiB
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML	86.2 MiB
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml	4.4 MiB
workflows	2 items
Workflow_Fib_N_O_Tryptic.toppas	64.7 KiB
Workflow_IgG_N_Tryptic.toppas	64.0 KiB

Figure 3: The extracted files from the zip folder within the example data set

The raw data can be analyzed within TOPPAS (see Section 2.1) using the files within the input directory and the corresponding workflow or the provided result files can be viewed/analyzed with the glyXtoolMS Evaluator (see Section 2.2).

2.1 New Analysis with TOPPAS

A) Start the glyXtool^{MS} Evaluator and edit the workflows like described in Section 1.4

B) Start TOPPAS

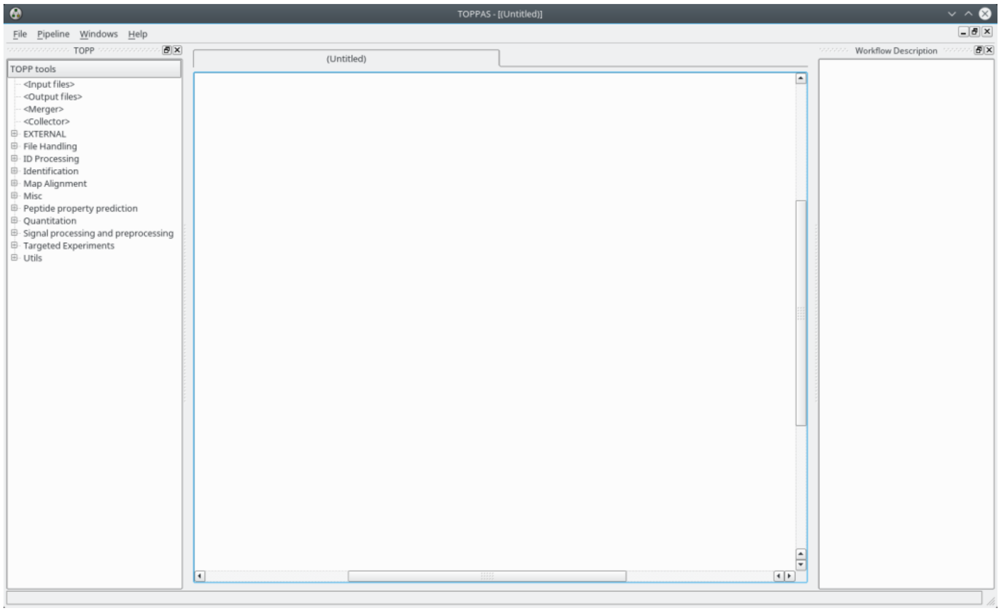


Figure 4: The TOPPAS window

C) Check the existence of the glyXtool^{MS} scripts. If tools are missing check Section 1.4

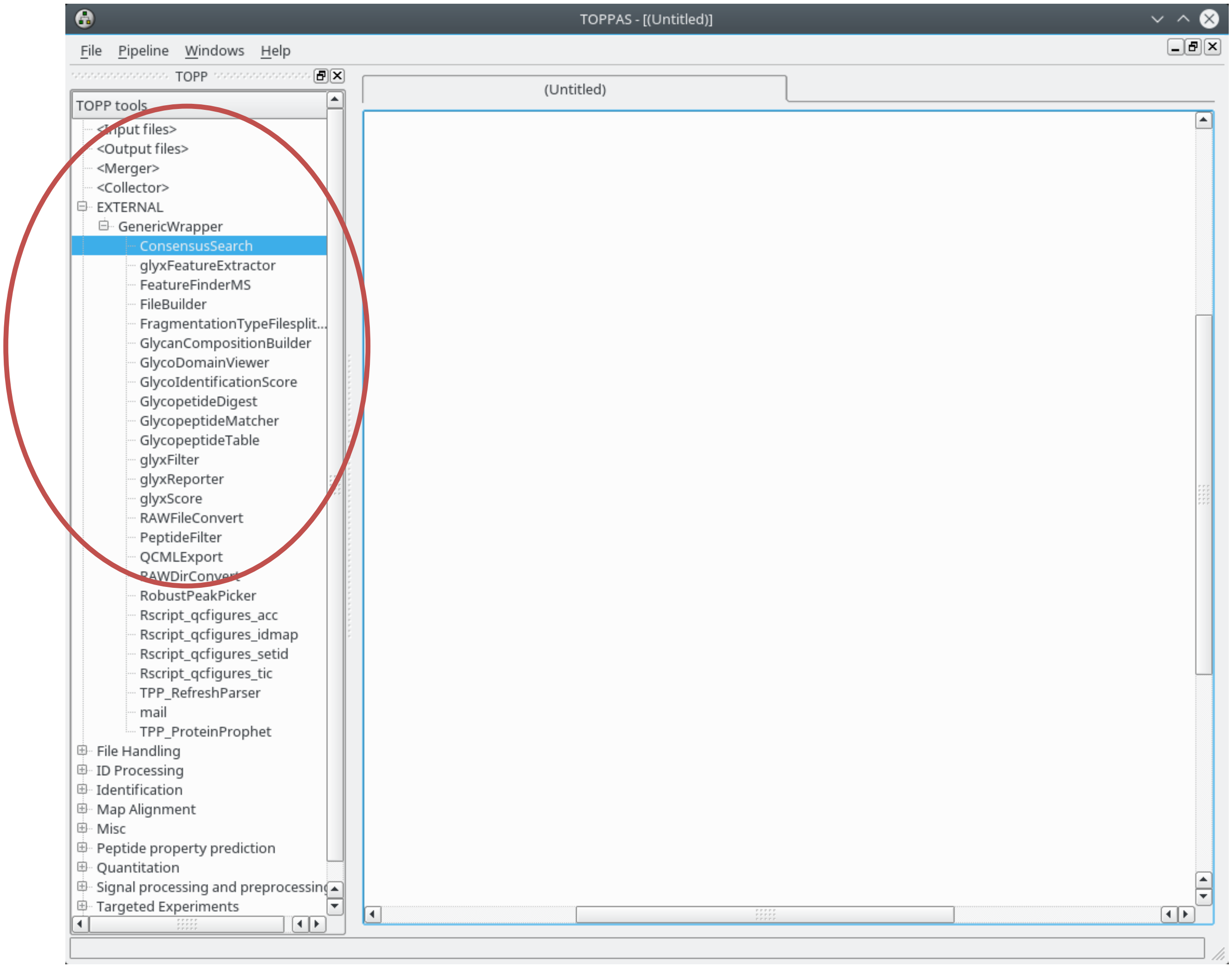


Figure 5: glyXtoolMS scripts are located under "External/GenericWrapper/…"

D) Open the IgG workflow

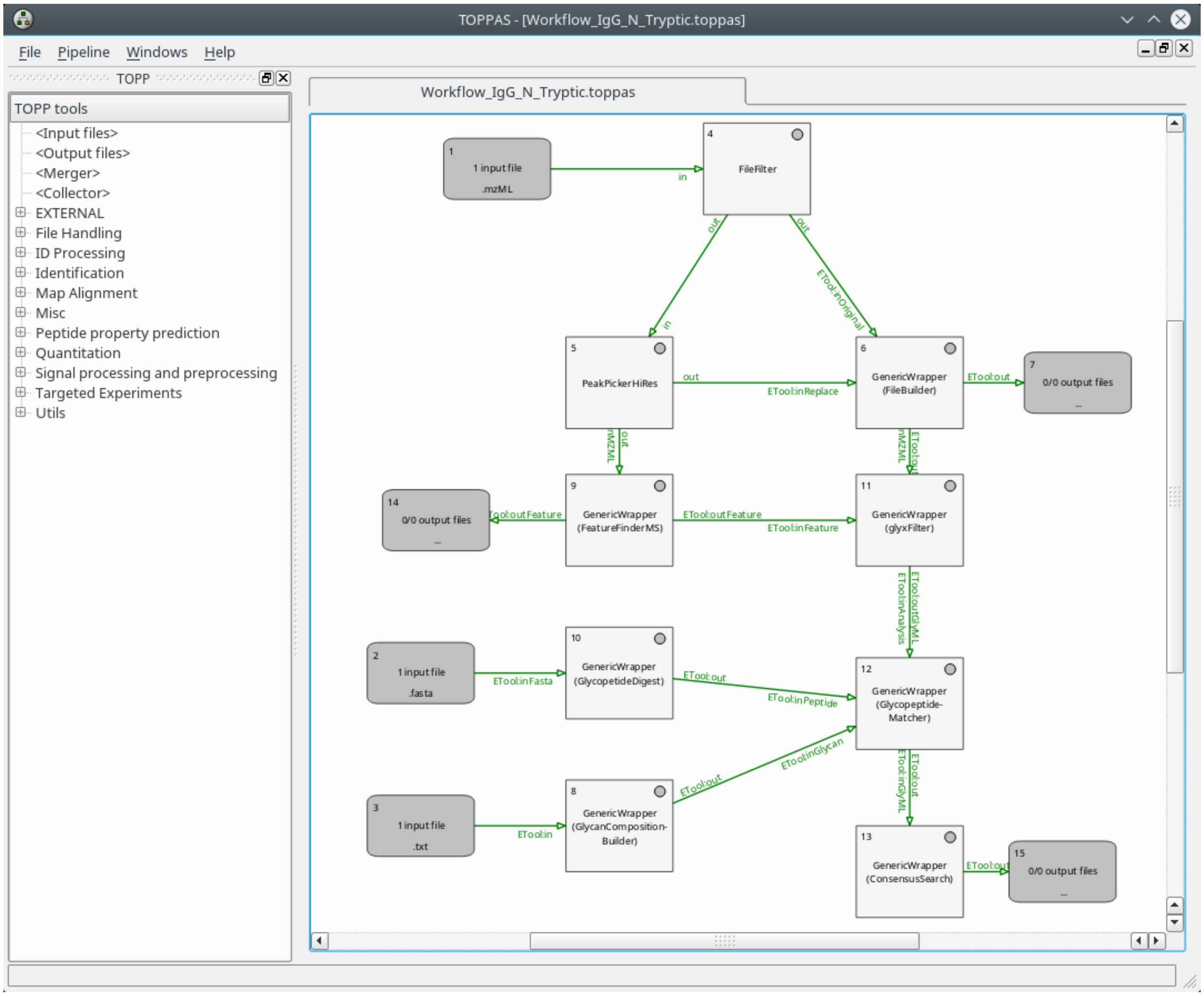
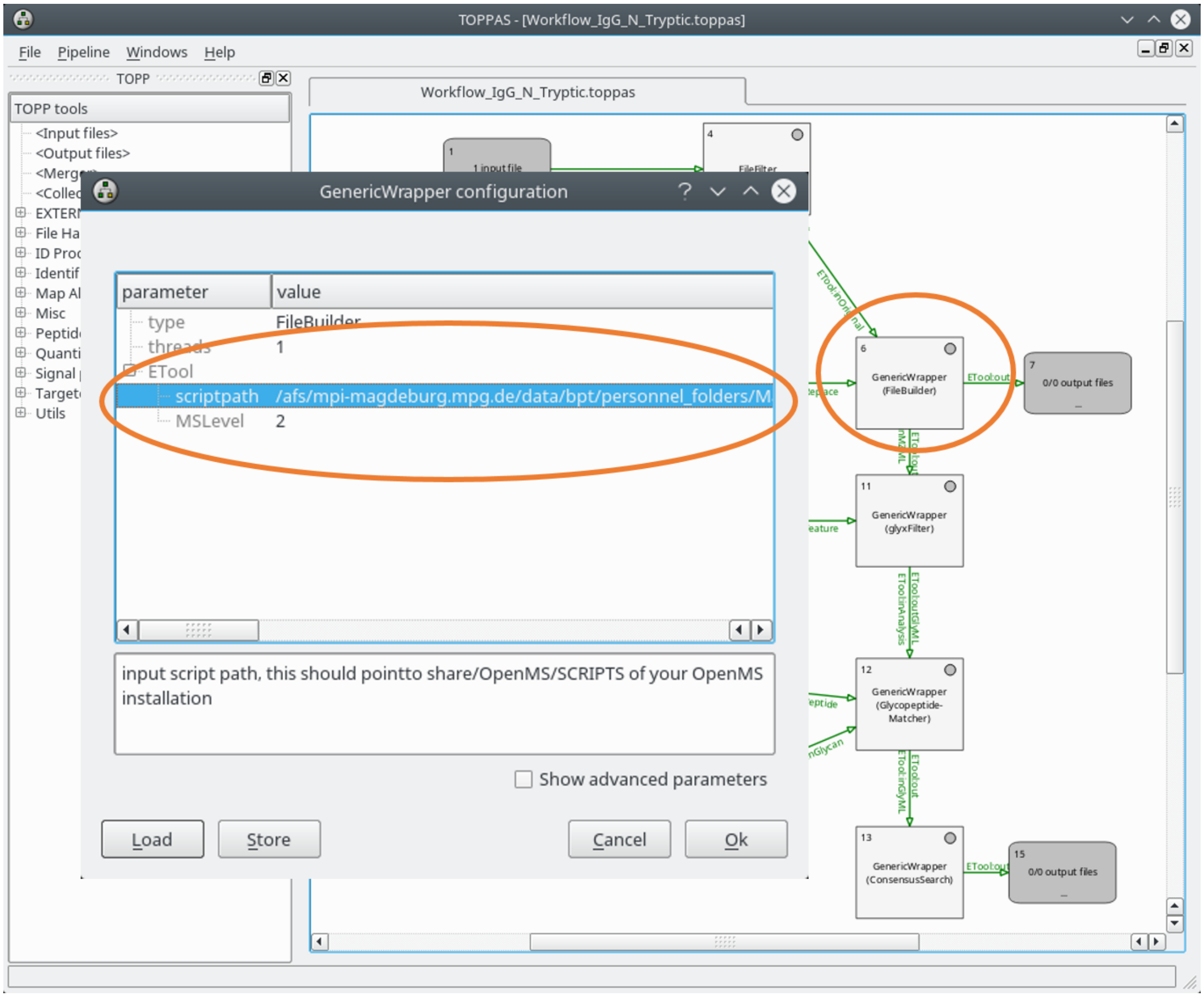


Figure 6: The IgG TOPPAS workflow

E) Double clicking on a tool opens the tool configuration. Check a Generic Wrapper tool for the correct OpenMS Script Path. If not correct, check Section 1.4.


TOPPAS - [Workflow_IgG_N_Tryptic.toppas]
File
Pipeline
Windows
Help
TOPP
TOPP tools
<Input files>
<Output files>
<Merge
<Collec
EXTERN
File Ha
ID Proc
Identif
Map Al
Misc
Peptid
Quanti
Signal
Target
Utils
Workflow_IgG_N_Tryptic.toppas
1 input file
FileFilter
GenericWrapper configuration
GenericWrapper (FileBuilder)
ETool:out
0/0 output files
GenericWrapper (glyxFilter)
GenericWrapper (Glycopeptide-Matcher)
GenericWrapper (ConsensusSearch)
Load
Store
Cancel
Ok

parameter	value
type	FileBuilder
threads	1
ETool	
scriptpath	/afs/mpi-magdeburg.mpg.de/data/bpt/personnel_folders/M
MSLevel	2

input script path, this should pointto share/OpenMS/SCRIPTS of your OpenMS installation

☐ Show advanced parameters

Figure 7: Check OpenMS Path

F) Set the input files:

- A: the IgG mass spectrometry file rawfiles/20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML
- B: the IgG FASTA file from input/FASTA Files/IgG_1_2_3_4.fasta
- C: the glycan database from input/Glycan DB/Human_N_and_O_glycans.txt

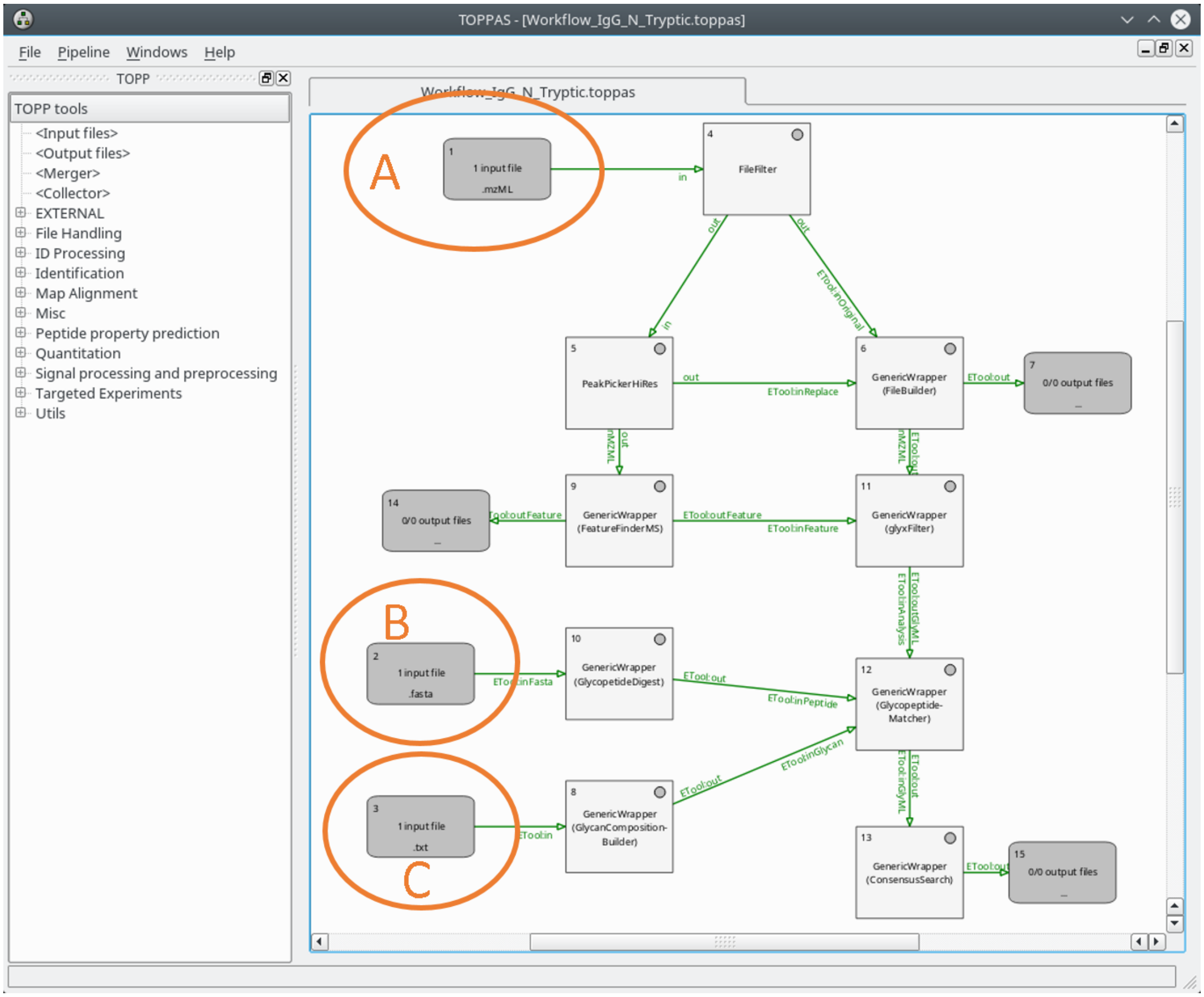
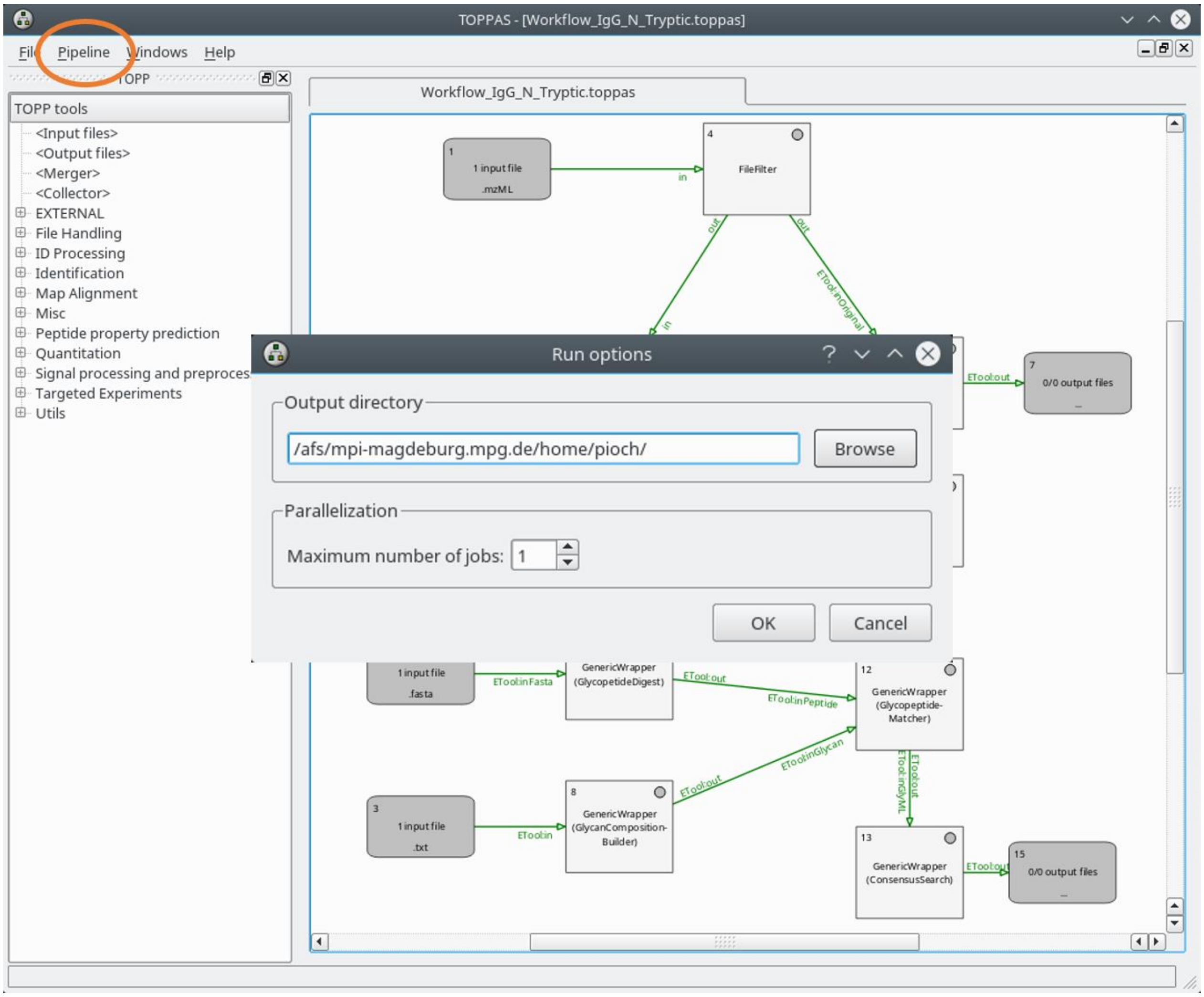


Figure 8: Set input files

G) Run the TOPPAS Pipeline. Select an output folder (TOPPAS will create a folder structure “TOPPAS_out” containing the files created by the output nodes)



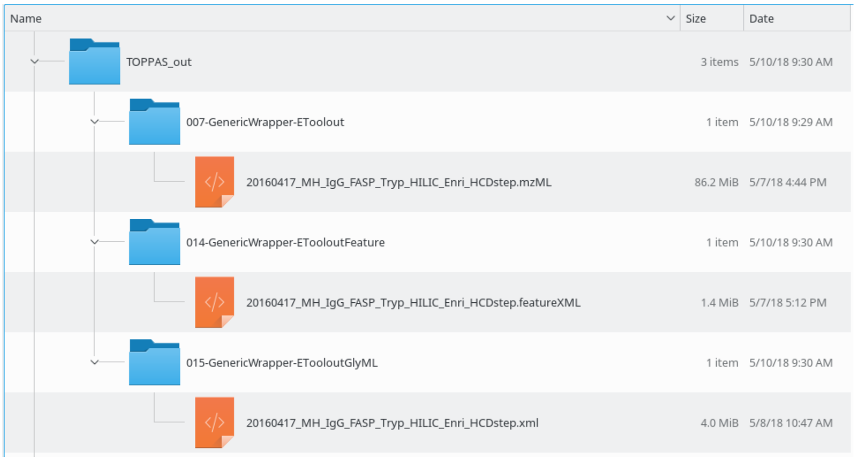
Output directory: /afs/mpi-magdeburg.mpg.de/home/pioch/

Parallelization

Maximum number of jobs: 1

Figure 9: Run the TOPPAS pipeline

H) The resulting analysis files are created during the successful run of the analysis:



Name	Size	Date
TOPPAS_out	3 items	5/10/18 9:30 AM
007-GenericWrapper-ETToolout	1 item	5/10/18 9:29 AM
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML	86.2 MiB	5/7/18 4:44 PM
014-GenericWrapper-ETooloutFeature	1 item	5/10/18 9:30 AM
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.featureXML	1.4 MiB	5/7/18 5:12 PM
015-GenericWrapper-ETooloutGlyML	1 item	5/10/18 9:30 AM
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml	4.0 MiB	5/8/18 10:47 AM

Figure 10: Analysis files created by the IgG TOPPAS workflow

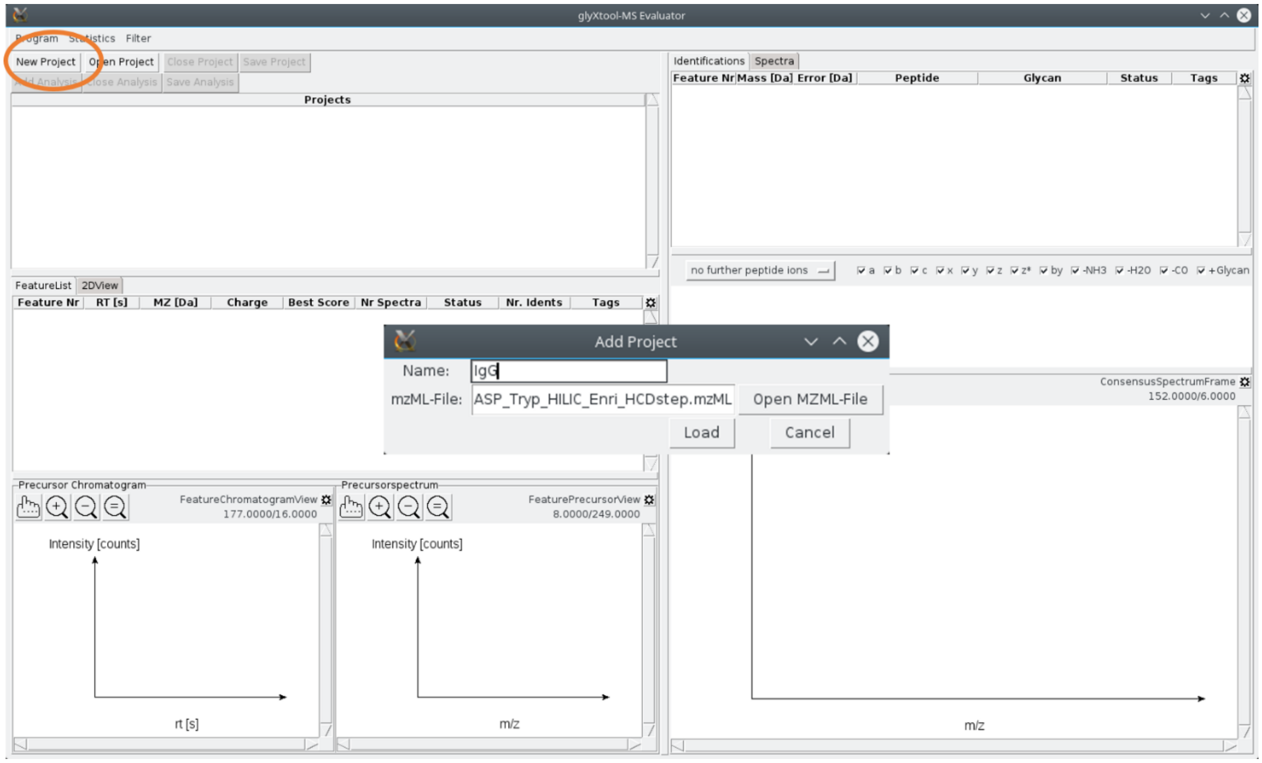
Move all files into a result folder. The created files should correspond to the ones within the example data set stored under “results/IgG”.

2.2 View Analysis with glyXtool^{MS} Evaluator

A) Start the glyXtool^{MS} Evaluator via commandline:

```
glyxtoolms
```

B) Create new Project with “New Project”. Select the mzML file from the result (it contains continuous MS1 spectra and centroided MS2 spectra) and provide a project name

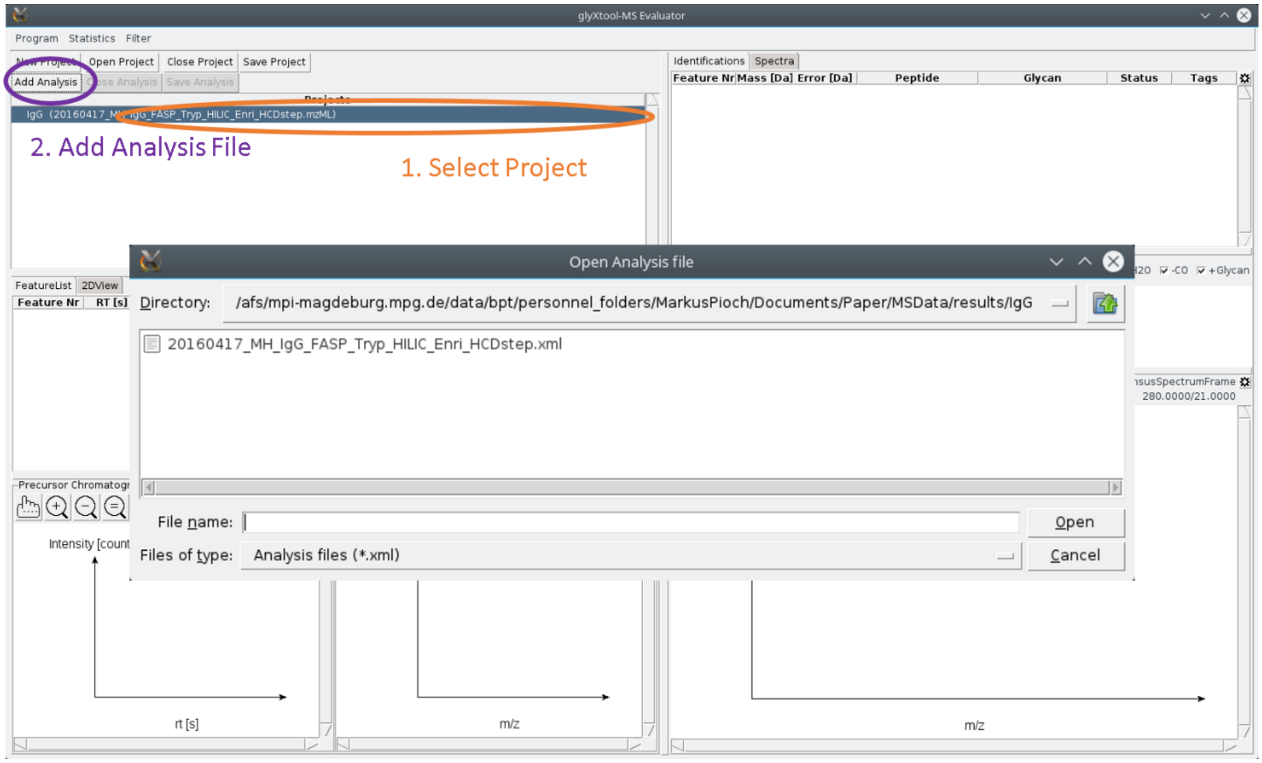


Name: IgG

mzML-File: ASP_Tryp_HILIC_Enri_HCDstep.mzML

Figure 11: Software Surface

C) Add an Analysis file to the project, by selecting the project, and then using the “Add Analysis Button”. To each Project multiple analysis files can be loaded (originating from the same raw data file). Saving the Project stores a simple file with the project name, its mzML file path and the path to each analysis file. This is only for easy access to the files – saving the project does not save changes to the analysis file!. Analysis files can be saved separately via the “Save Analysis” Button.



IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)

Directory: /afs/mpi-magdeburg.mpg.de/data/bpt/personnel_folders/MarkusPioch/Documents/Paper/MSData/results/IgG

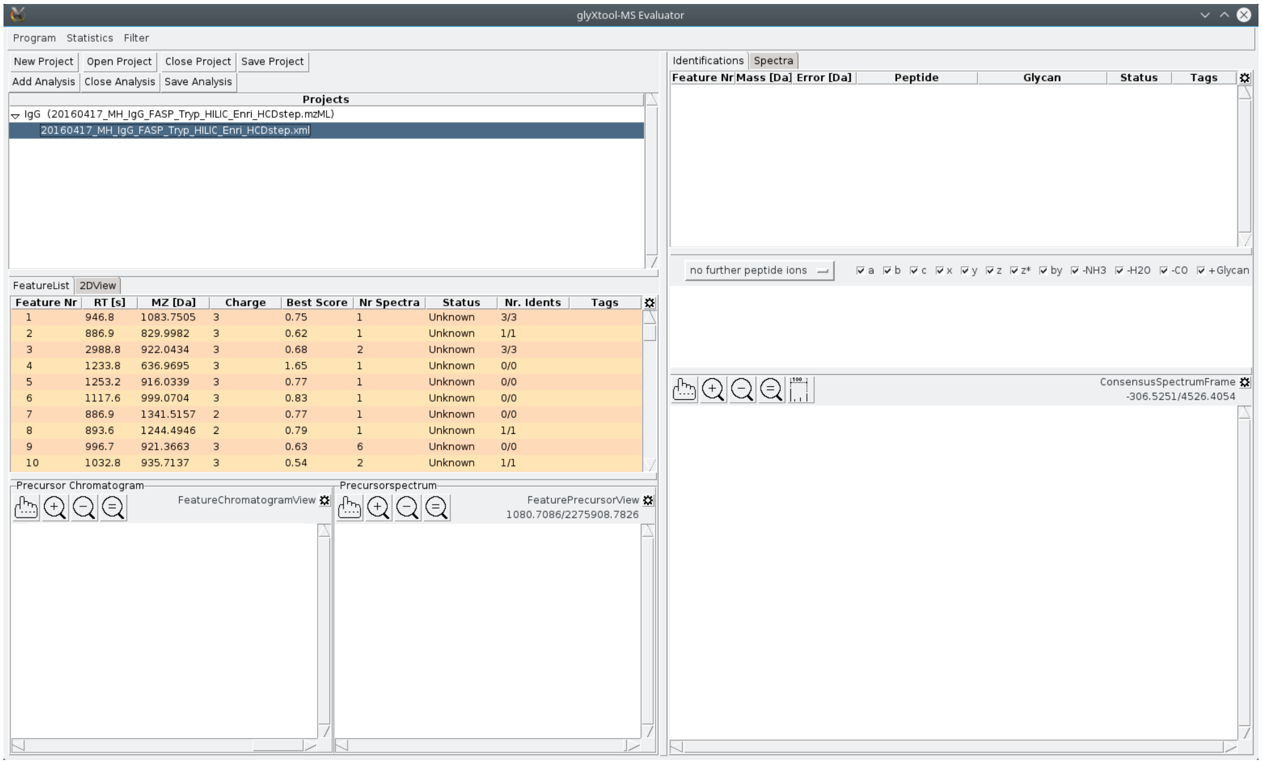
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

File name:

Files of type: Analysis files (*.xml)

Figure 12: Add Analysis

D) Loaded Analysis File: In the “FeatureList” all features/compounds are listed that were found by the FeatureFinder and with have been identified as potential glycoepetides via the glyXFilter tool. The best Oxonoimion-Score is shown within the table.



Projects

- IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)
  - 20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

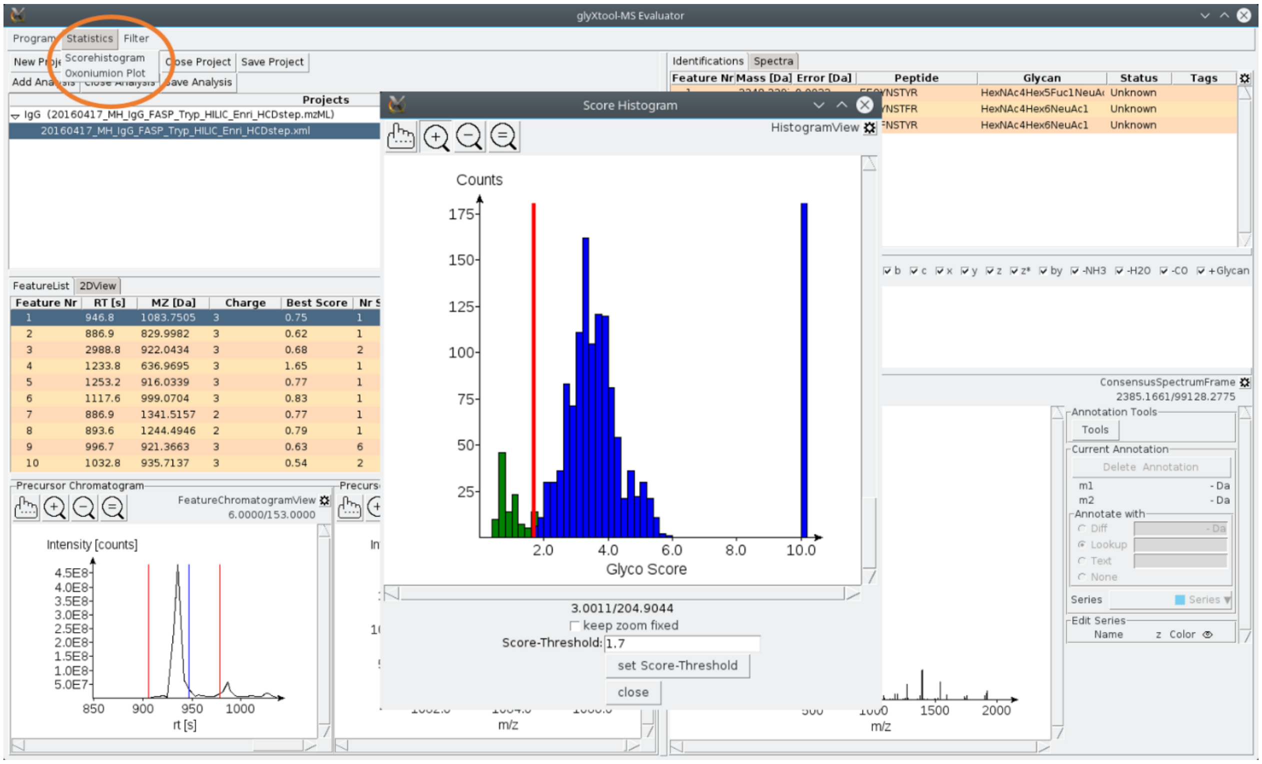
Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags

no further peptide ions ☑ a ☑ b ☑ c ☑ x ☑ y ☑ z ☑ z* ☑ by ☑ -NH3 ☑ -H2O ☑ -CO ☑ +Glycan

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Unknown	3/3	
2	886.9	829.9982	3	0.62	1	Unknown	1/1	
3	2988.8	922.0434	3	0.68	2	Unknown	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

Figure 13: Loaded Analysis file

E) Show the oxonium scoring results for the glyXFilter tool:



IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr S
1	946.8	1083.7505	3	0.75	1
2	886.9	829.9982	3	0.62	1
3	2988.8	922.0434	3	0.68	2
4	1233.8	636.9695	3	1.65	1
5	1253.2	916.0339	3	0.77	1
6	1117.6	999.0704	3	0.83	1
7	886.9	1341.5157	2	0.77	1
8	893.6	1244.4946	2	0.79	1
9	996.7	921.3663	3	0.63	6
10	1032.8	935.7137	3	0.54	2

Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
1	2340.330	0.0022	EEQYNSTYR	HexNAc4Hex5Fuc1NeuAc	Unknown	
			YNSTFR	HexNAc4Hex6NeuAc1	Unknown	
			FNSTYR	HexNAc4Hex6NeuAc1	Unknown	

☑ b ☑ c ☑ x ☑ y ☑ z ☑ z* ☑ by ☑ -NH3 ☑ -H2O ☑ -CO ☑ +Glycan

m1 - Da
m2 - Da

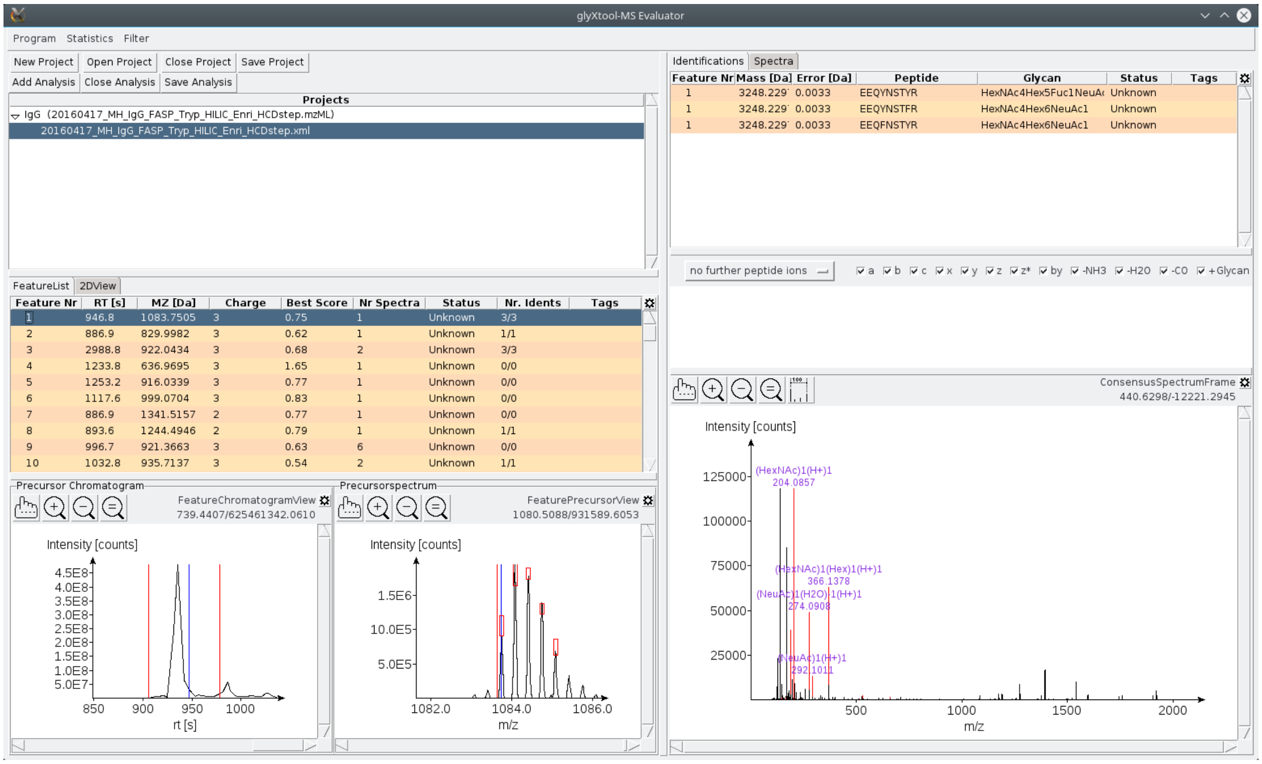
○ Diff - Da
● Lookup
○ Text
○ None

Series 🟦 Series

3.0011/204.9044
□ keep zoom fixed
Score-Threshold: 1.7
set Score-Threshold
close

Figure 14: Histogram of Oxonium ion scoring. The threshold should divide two distinct populations of glycopeptids (green) and non glycopeptides (blue)

F) Selecting a Feature within the “FeatureList” shows the extracted ion chromatogram of the monoisotopic precursor peak, the isotopic pattern of the precursor and the consensus spectrum of all fragment spectra within its feature box. On the right side within the Identifications tab – all possible precursor mass matches of theoretical peptides and glycan compositions are shown, generated by the GlycopeptideMatcher tool. The Spectra tab shows the single fragment spectra associated with the feature.



glyXtool-MS Evaluator

Program Statistics Filter

New Project | Open Project | Close Project | Save Project

Add Analysis | Close Analysis | Save Analysis

Projects

IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)

20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

FeatureList | 2DView

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Unknown	3/3	
2	886.9	829.9982	3	0.62	1	Unknown	1/1	
3	2988.8	922.0434	3	0.68	2	Unknown	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

Identifications | Spectra

Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
1	3248.229	0.0033	EEQYNSTYR	HexNAc4Hex5Fuc1NeuAc	Unknown	
1	3248.229	0.0033	EEQYNSTFR	HexNAc4Hex6NeuAc1	Unknown	
1	3248.229	0.0033	EEQFNSTYR	HexNAc4Hex6NeuAc1	Unknown	

no further peptide ions ☑ a ☑ b ☑ c ☑ x ☑ y ☑ z ☑ z* ☑ by ☑ -NH3 ☑ -H2O ☑ -CO ☑ +Glycan

Precursor Chromatogram

FeatureChromatogramView 739.4407/625461342.0610

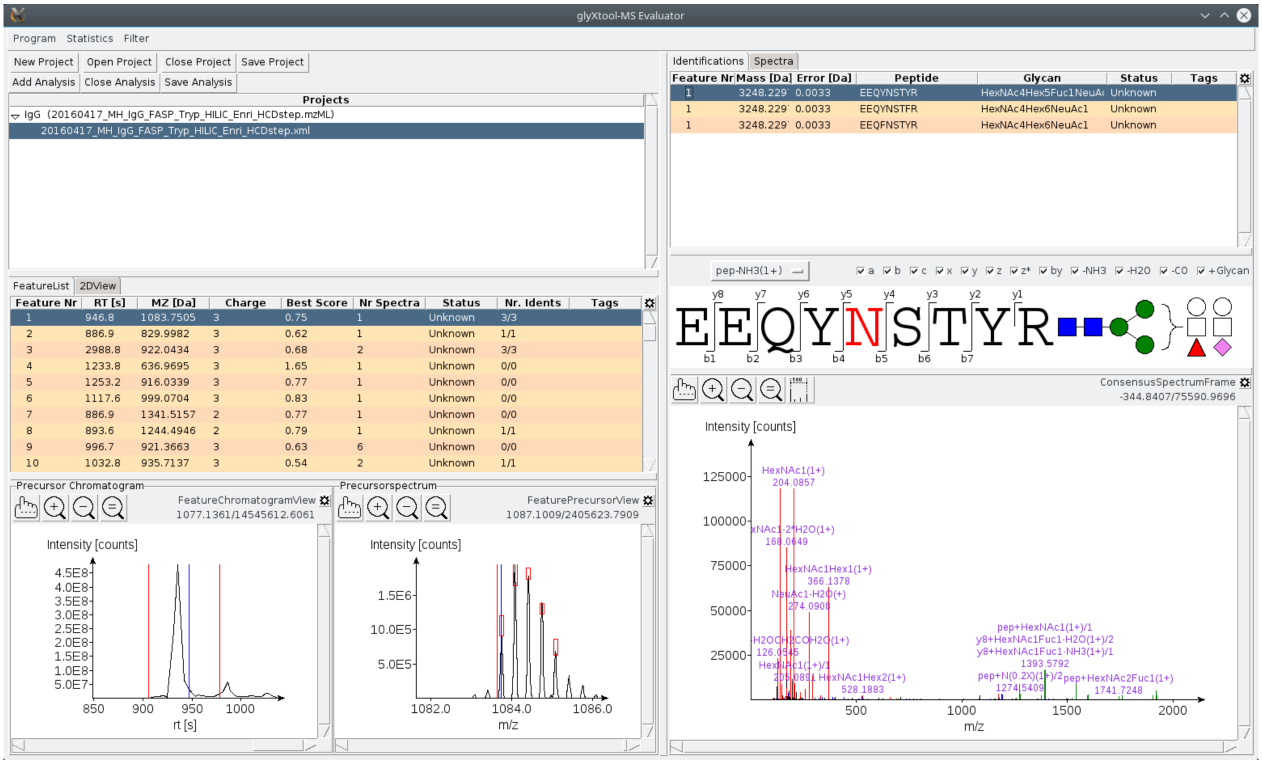
Precursorspectrum

FeaturePrecursorView 1080.5088/931589.6053

ConsensusSpectrumFrame 440.6298/-12221.2945

Figure 15: Selecting a feature

G) Selecting an identification within the “Identifications” tab annotates the consensus spectrum with its theoretical ion fragments. Using the “Gear icon” a configuration panel for the consensus spectrum is shown, where ion names and colors can be adapted.



Projects

- IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)
  - 20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Unknown	3/3	
2	886.9	829.9982	3	0.62	1	Unknown	1/1	
3	2988.8	922.0434	3	0.68	2	Unknown	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

FeatureChromatogramView
1077.1361/14545612.6061

FeaturePrecursorView
1087.1009/2405623.7909

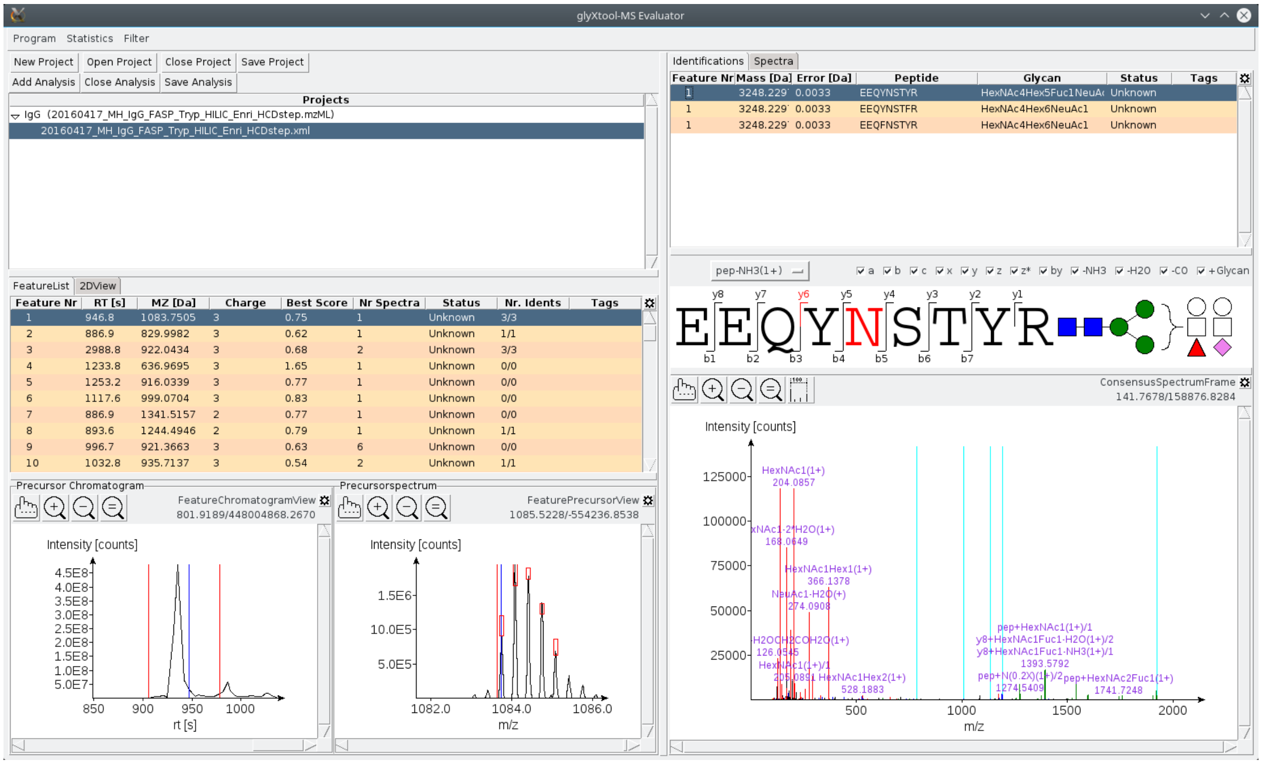
Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
1	3248.229	0.0033	EEQYNSTYR	HexNAc4Hex5Fuc1NeuA	Unknown	
1	3248.229	0.0033	EEQYNSTFR	HexNAc4Hex6NeuAc1	Unknown	
1	3248.229	0.0033	EEQFNSTYR	HexNAc4Hex6NeuAc1	Unknown	

pep-NH3(1+) ☑ a ☑ b ☑ c ☑ x ☑ y ☑ z ☑ z* ☑ by ☑ -NH3 ☑ -H2O ☑ -CO ☑ +Glycan

ConsensusSpectrumFrame
-344.8407/75590.9696

Figure 16: Selecting an identification

H) Selecting a peptide ion shows all occurrences of this ion with the fragment spectrum



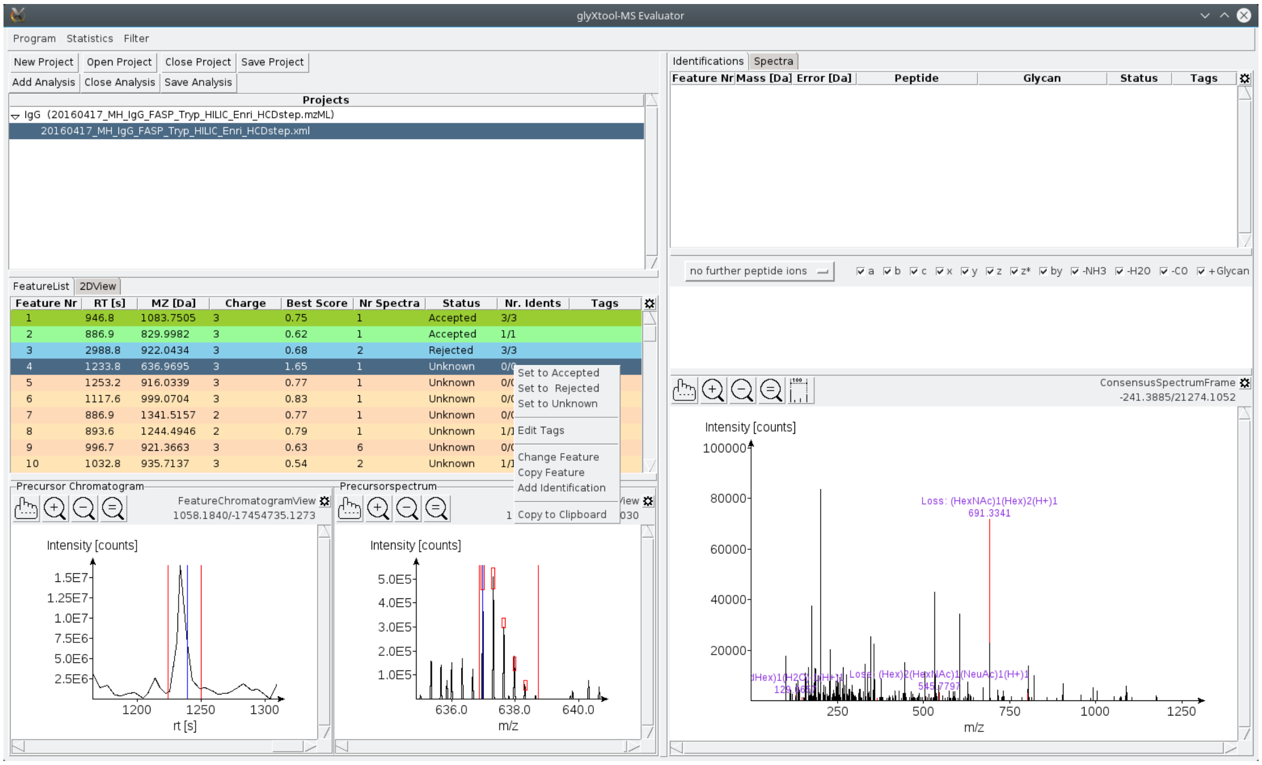
Projects
IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Unknown	3/3	
2	886.9	829.9982	3	0.62	1	Unknown	1/1	
3	2988.8	922.0434	3	0.68	2	Unknown	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
1	3248.229'	0.0033	EEQYNSTYR	HexNAc4Hex5Fuc1NeuAc	Unknown	
1	3248.229'	0.0033	EEQYNSTFR	HexNAc4Hex6NeuAc1	Unknown	
1	3248.229'	0.0033	EEQFNSTYR	HexNAc4Hex6NeuAc1	Unknown	

Figure 17: Selecting a peptide ion

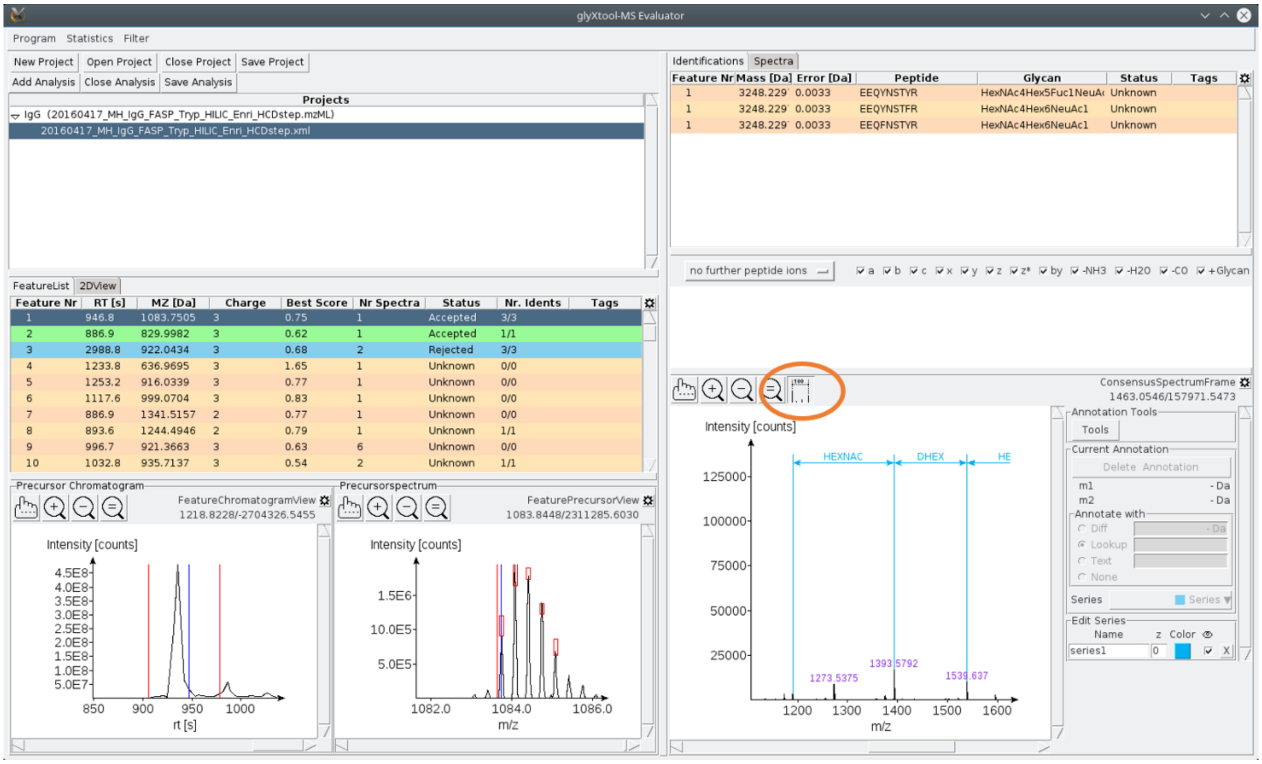
I) Right clicking on a selected feature or identification opens a context menu for analysis. A Status of “Unknown”, “Accepted” or “Rejected” can be set, tags can be added, etc.



Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Accepted	3/3	
2	886.9	829.9982	3	0.62	1	Accepted	1/1	
3	2988.8	922.0434	3	0.68	2	Rejected	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/(	
5	1253.2	916.0339	3	0.77	1	Unknown	0/(	
6	1117.6	999.0704	3	0.83	1	Unknown	0/(	
7	886.9	1341.5157	2	0.77	1	Unknown	0/(	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/(	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

Figure 18: Context menu on either features or identifications

J) Fragment ion annotation series can be added by activating the ruler icon: Either by left clicking on the peak of interest and pulling to the side, or right clicking where a menu shows potentially interesting mass differences to neighbor peaks.



Projects

IgG (20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.mzML)
20160417_MH_IgG_FASP_Tryp_HILIC_Enri_HCDstep.xml

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Accepted	3/3	
2	886.9	829.9982	3	0.62	1	Accepted	1/1	
3	2988.8	922.0434	3	0.68	2	Rejected	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

FeatureChromatogramView 1218.8228/-2704326.5455

FeaturePrecursorView 1083.8448/2311285.6030

Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
1	3248.229	0.0033	EEQYNSTYR	HexNAc4Hex5Fuc1NeuAc	Unknown	
1	3248.229	0.0033	EEQYNSTFR	HexNAc4Hex6NeuAc1	Unknown	
1	3248.229	0.0033	EEQFNSTYR	HexNAc4Hex6NeuAc1	Unknown	

no further peptide ions ☑ a ☑ b ☑ c ☑ x ☑ y ☑ z ☑ z* ☑ by ☑ -NH3 ☑ -H2O ☑ -CO ☑ +Glycan

ConsensusSpectrumFrame 1463.0546/157971.5473

m1 - Da
m2 - Da

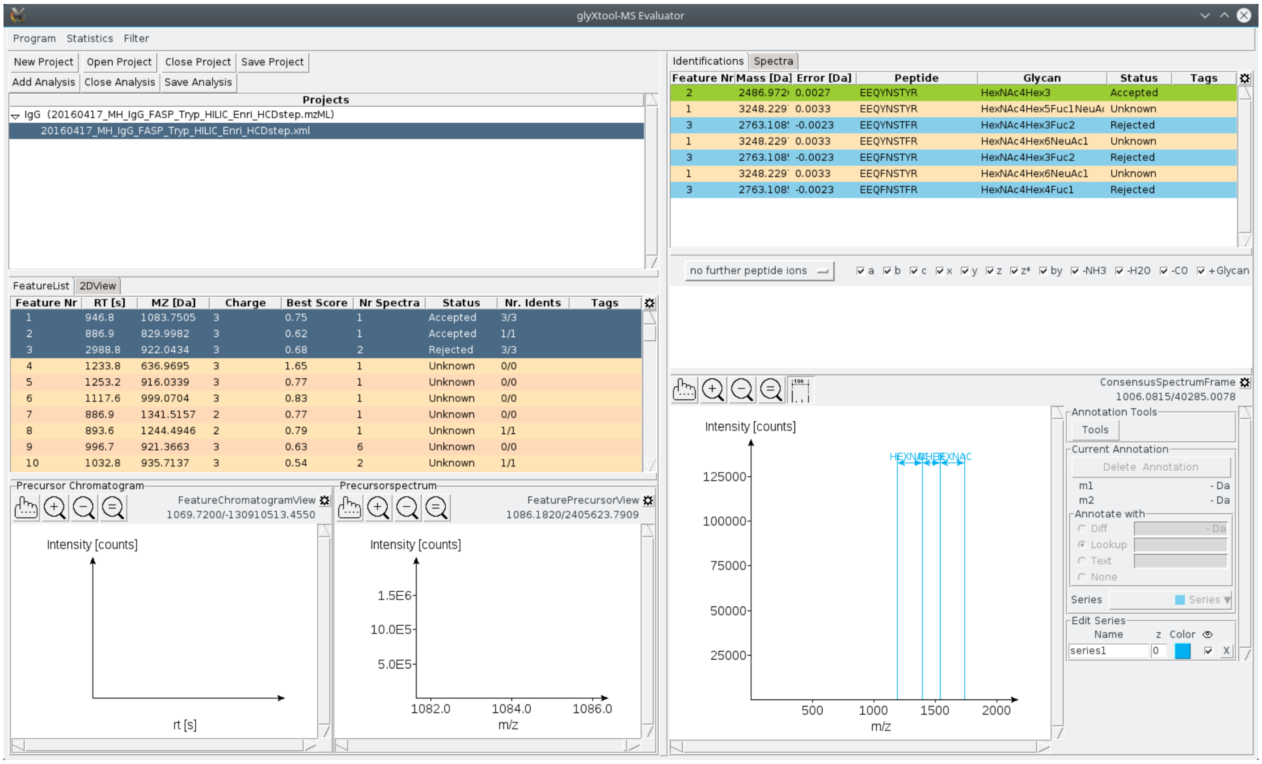
○ Diff - Da
● Lookup
○ Text
○ None

Series 🟦 Series

Name	z	Color	👁	
series1	0	🟦	☑	X

Figure 19: Adding fragment ions annotation series

K) Both the “FeatureList” and the “Identifications” tab support multi selection, if e.g. two features are selected, the identifications are shown for all. By selecting all features with “Cntrl+A” all identifications can be selected, for example.



Feature Nr	Mass [Da]	Error [Da]	Peptide	Glycan	Status	Tags
2	2486.972	0.0027	EEQYNSTYR	HexNAc4Hex3	Accepted	
1	3248.229	0.0033	EEQYNSTYR	HexNAc4Hex5Fuc1NeuA	Unknown	
3	2763.108	-0.0023	EEQYNSTFR	HexNAc4Hex3Fuc2	Rejected	
1	3248.229	0.0033	EEQYNSTFR	HexNAc4Hex6NeuAc1	Unknown	
3	2763.108	-0.0023	EEQFNSTYR	HexNAc4Hex3Fuc2	Rejected	
1	3248.229	0.0033	EEQFNSTYR	HexNAc4Hex6NeuAc1	Unknown	
3	2763.108	-0.0023	EEQFNSTFR	HexNAc4Hex4Fuc1	Rejected	

Feature Nr	RT [s]	MZ [Da]	Charge	Best Score	Nr Spectra	Status	Nr. Idents	Tags
1	946.8	1083.7505	3	0.75	1	Accepted	3/3	
2	886.9	829.9982	3	0.62	1	Accepted	1/1	
3	2988.8	922.0434	3	0.68	2	Rejected	3/3	
4	1233.8	636.9695	3	1.65	1	Unknown	0/0	
5	1253.2	916.0339	3	0.77	1	Unknown	0/0	
6	1117.6	999.0704	3	0.83	1	Unknown	0/0	
7	886.9	1341.5157	2	0.77	1	Unknown	0/0	
8	893.6	1244.4946	2	0.79	1	Unknown	1/1	
9	996.7	921.3663	3	0.63	6	Unknown	0/0	
10	1032.8	935.7137	3	0.54	2	Unknown	1/1	

Figure 20: Multiselection of several features displays all corrsponsing identifications within the “Identifications” tab

3 TOPPAS tools for glycopeptide analytics

Here, the purpose of each tool and its possible predecessor tools and successor tools are described.

3.1 FeatureFinderMS

pot. predecessor tools		pot. successor tools
MS centroid data in .mzML format PeakPickerHighRes PeakPickerWavelet FileFilter	→ FeatureFinderMS →	glyxFilter

Purpose

Finds features around analytes containing at least one fragment spectrum.

Parameters

- inMZML: Input mass spectra as centroid data in *.mzML file format
- outFeature: Feature output file
- tolerance: Mass tolerance in Dalton
- mswindow: maximum mass range of the precursor isotope pattern in dalton
- precursorshift: maximum deviation of the precursor mass from the (average) precursor mass reported within the mass file in dalton
- rtwindow: maximum elution range of the analyte peak in seconds

Possible Input Nodes

The tool uses centroided MS1 data. Possible input nodes are the file input node, the various OpenMS Peakpicker nodes or the FileFilter node if the data have to be cropped to a certain elution or mass range

Possible Output Nodes

- glyxFilter

Similar tools

- FeatureFinderCentroided
- FeatureFinderisotopeWavelet

3.2 FileBuilder

pot. predecessor tools		pot. successor tools
FileFilter PeakPicker	→ FileBuilder →	glyxFilter
FileFilter PeakPicker		

Purpose

Replaces the given MS Level spectra in an experiment. In the context of glycopeptide analysis it is used to replace continuous MS^2 fragment spectra with their centroided counterpart after peakpicking, while retaining continuous data in the MS^1 domain. This is needed as input for the 'glyXtool^{MS} Evaluator' to visualize continuous MS1 data for the precursors.

Parameters

- inOriginal: File input of mass spectrometry data in *.mzML format; All MS level are transferred to the output file except the level provided by the option 'MSLevel'
- inReplace: File input of mass spectrometry data in *.mzML format; The spectra matching the given MSLevel option are transferred to the output file
- out: File output in *.mzML format
- MSLevel: MS level which will be replaced with data from the replacement file

Possible Input Nodes

- FileFilter
- PeakPicker
- File input node

Possible Output Nodes

- glyXfiler
- File output node

3.3 glyXFilter

pot. predecessor tools		pot. successor tools
FeatureFinderCentroided FeatureFinderMS* FeatureFinderIsotopeWavelt	→ glyxFilter →	glyxReporter Glycopeptide Matcher* File Output node (.xml)
MS centroid data in .mzML format PeakPickerHighRes PeakPickerWavelet FileFilter FileBuilder*		

Purpose

The tool searches for glycopeptide evidence in MS^2 spectra, based on oxonium ions and neutral losses from the precursor. Reported is a spectrum score between 0.0 and 10.0 for each MS2 spectrum where the lower score signifies a higher glycopeptide probability. The identified glycopeptide fragment spectra are then used to identify glycopeptide features in the FeatureMap. For easier data access in later stages of the analysis pipeline the tool then generates consenus spectra for all identified glycopeptide features. All generated information is finally stored in a *.xml file.

Parameters

- inMZML: Input mass spectra as centroid data in *.mzML file format
- inFeature: Input feature file as*.featureXML
- outGlyML: Output file in *.xml format, containing all scored fragment spectra and all identified glycopeptide features
- createFeatures: (false/true);when no feature could be found within the provided feature map for a given fragment spectra a dummy feature will be generated, if set to true
- hasFucose: (false/true); if true use predefined oxonium ions that contain fucose
- hasNANA: (false/true); if true use predefined oxonium ions that contain N-acetylneuraminic acid
- hasNGNA: (false/true); if true use predefined oxonium ions that contain N-glycolylneuraminic acid
- oxoniumions: Add additional oxonium ions to the search.
  Format has to be like: (NeuAc)1(H2O)-1(H+)1 with comma separated oxonium ions
- tolerance: Mass tolerance for the oxonium ion search
- toleranceType: Mass tolerance Type (either ppm or Da)
- ionthreshold: Ignores peaks with lower intensity than the given threshold. Set to 0 to include all peaks.
- scorethreshold: Threshold used to identify a fragment spectrum as a glycopeptide. Lower scores signify a higher glycopeptide probability.

Possible Input Nodes

- inMZML: needed are centroided MS2 data with sorted peaks after mass. Suitable is the FileBuilder, to generate a suitable input file for the glyXtool Evaulator; PeakPicker or FileFilter if the MS2 data are already centroid data and need to be sorted
- inFeature: All possible FeatureFinder tools

Possible Output Nodes

- glyxReporter:
- glycoPeptideMatcher: for matching peptide and glycan composition to the precursor masses of identified glycopeptide features

3.4 GlycopeptideDigest

pot. predecessor tools		pot. successor tools
Protein sequences in *.fasta format	→ Glycopeptide Digest →	GlycopeptideMatcher

Purpose

Generating possible peptide sequences with glycosylation sites from protein sequences via theoretical digest.

Parameters

- inFasta: Input file in *.fasta format containing either protein sequences or peptide squences
- out: *.xml file containing the generated peptides with glycosylation sites, their possible modifications and the monoisotopic mass of each peptide
- enzymes: The enzyme(s) used for the digest. Currently supported are trypsin, trypsin/P, AspN, Unspecific and NoDigest. The option 'Unspecific' cuts after each aminoacid and uses the Nr of missedCleavageSites as the maximum length of the reported peptides. With the option 'NoDigest' the provided sequences from the *.fasta file are used without digest, allowing the user to specify peptides.
- maxNrModifications: Nr of maximum allowed modifications on each peptide. CYS_CAM and CYS_CM are excluded.
- modifications: Variable modifications. For each peptide all possible permutations are generated. If e.g. a peptide contains two methionines, three peptides are generated: (0 Oxidations, 1 Oxidation on either methionine and fully oxidized on both residues)
- glycosylation: (N-glycosylation, O-glycosylation). Select which glycosylation site should be checked. Uses the motif N(S|T)(^P) as consensus sequence for N-glycosylation and (S|T) for O-glycosylation
- missedCleavageSites: maximum nr of missed cleavage sites. In case of unspecific digest determines the maximum length of the peptide.

Possible Input Nodes

- Input node with *.fasta file

Possible Output Nodes

- GlycopeptideMatcher

3.5 GlycanComposition builder

pot. predecessor tools		pot. successor tools
File with glycan compositions Or no file node	→ Glycan Composition Builder →	GlycopeptideMatcher

Purpose

Provides glycan compositions for the ‘Glycopeptide Matcher’ tool. A given list of glycan compositions can be filtered by the provided ranges if the ‘useAsFilter’ option is set to true, otherwise a list of glycan compositions is calculated in-silico with the given ranges.

Parameters

- in: File input
- out: Output file, containing the filtered glycan compostions in an *.txt file
- useAsFilter: (false/true); If true, filters the glycan compositions from the input file according to the provided monomer ranges. If false disregards content of the input file and calculates all compositon permutations from the given monomer ranges.
- rangeHex: range of hexose within the glycan composition
- rangeHexNAc: range of N-acetylhexoseamine within the glycan composition
- rangeFuc: range of fucose within the glycan composition
- rangeNeuAc: range of N-acetylneuraminic acid within the glycan composition
- rangeNeuGc: range of N-glycolylneuraminic acid within the glycan composition

Possible Input Nodes

- File input node. In case the ‘useAsFilter’ is set to false, the content of the file input node is neglected, since the glycan compostion permutations are calculated based on the given ranges. However TOPPAS expects each tool to have an input node to run, thus some file has to be provided to the tool.

Possible Output Nodes

- Glycopeptide Matcher

3.6 Glycopeptide Matcher

pot. predecessor tools		pot. successor tools
glyxFilter*	→ Glycopeptide Matcher →	glyxReporter* ConsensusSearch File Output node (.xml)
Glycopeptide Digest*		
Glycan Composition Builder File with glycan compositions		

Purpose

Matches a given list of peptides and glycan compositions to precursor masses of glycopeptide features.

Parameters

- out: Output file, Appends new collected information to the given inAnalysis file
- inAnalysis: Input file containing a glyML analysis file with scored glycopeptide features
- inGlycan: Input file containing a list of glycan compositions to match against
- inPeptide: Input file containing a list of peptide to match against
- tolerance: Mass tolerance for the oxonium ion search
- toleranceType: Mass tolerance Type (either ppm or Da)

Possible Input Nodes

a) glyxFilter
b) GlycopeptideDigest
c) Glycan Composition Builder
File with glycan compositions

Possible Output Nodes

- glyxReporter
- ConsensusSearch

3.7 Consensus Search

pot. predecessor tools		pot. successor tools
Glycopeptide Matcher	→ Consensus Search →	glyxReporter* File Output node (.xml)

Purpose

Annotates the consensus spectra of glycopeptide features with peptide fragments based on the theoretical fragments of the peptide sequence suggested by the Glycopeptide Matcher tool.

Parameters

- inGlyML: Input analysis file in glyML format
- outGlyML: Output analysis file in glyML format
- tolerance: Mass tolerance for the oxonium ion search
- toleranceType: Mass tolerance Type (either ppm or Da)
- ionthreshold: Intensity threshold for annotating fragment spectra peaks. Set to Zero to ignore intensity.
- pepIons: List of peptide ions to search for

Possible Input Nodes

- Glycopeptide Matcher

Possible Output Nodes

- glyxReporter

3.8 glyxReporter

pot. predecessor tools		pot. successor tools
glyxFilter Glycopeptide Matcher Fragment Search	→ glyxReporter →	File Output node (.xls)

Purpose

Converts the collected information stored in the glyML file from the glycopeptide analysis tools into excel sheets.

Parameters

- inAnalysis: Input file in glyML format
- outReport: Output as *.xls file

Possible Input Nodes

- glyxFilter
- Glycopeptide Matcher
- Fragment Search

Possible Output Nodes

- File output (*.xls)