

SugarQb

HOWTO, example workflow and data files.

(Version 5.9.2018)

Introduction:

SugarQb (www.imba.oeaw.ac.at/sugarqb) is a freely available collection of computational tools for the automated identification of intact glycopeptides from high-resolution HCD MS/MS data-sets in the Proteome Discoverer environment. SugarQb has now been migrated to the latest and free version of Proteome Discoverer 2.1

For further information on the algorithm, please refer to the corresponding publication Stadlmann. J., Taubenschmid J. Mechtler K., Penninger JM., et al. Comparative glycoproteomics of stem cells identifies new players in ricin toxicity, Nature (2017).

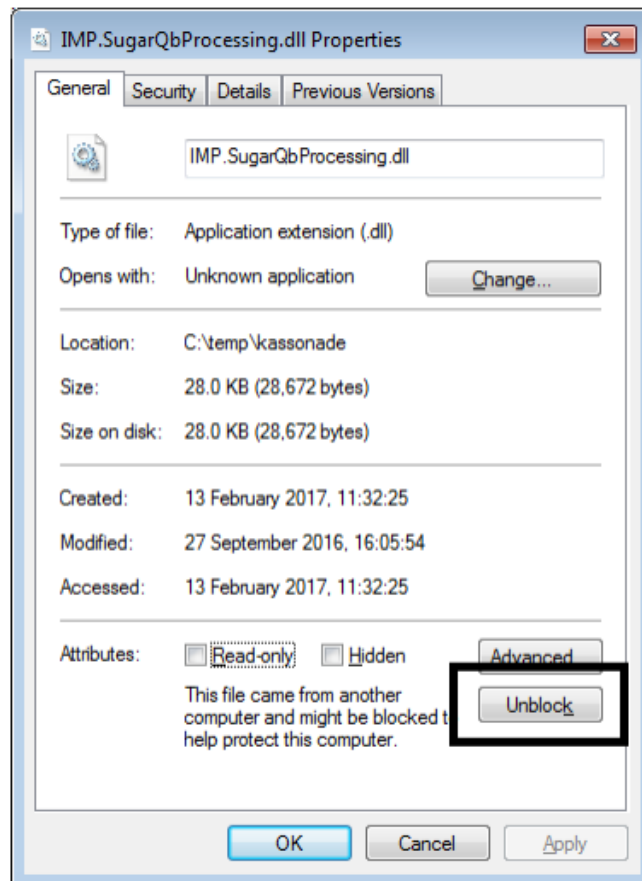
This document is intended to provide you with a quick guide on how to download, install and test SugarQb, analyzing example data of tryptic glycopeptides derived from human plasma. All relevant .dll files, additional parameter files and a Glycan mass data-base are available at:

www.imba.oeaw.ac.at/SugarQb.

Contact: SugarQb@imp.oeaw.ac.at

Download and Installation:

- Download SugarQb for Thermo Scientific Proteome Discoverer 2.1 using the following URL: www.imba.oeaw.ac.at/SugarQb
- Save all your files and shutdown Thermo Scientific Proteome Discoverer
- Navigate to the folder where you have installed Thermo Scientific Proteome Discoverer (Tip: You can easily find out the path by right-clicking the Thermo Scientific Proteome Discoverer desktop icon and open the Properties window. The older path is written in the field Target.)
- Copy the .dll files into the Thermo Scientific Proteome Discoverer folder.
- Unblock the .dll files if required by right-clicking each .dll file, opening its properties window and clicking the Unblock button if available.



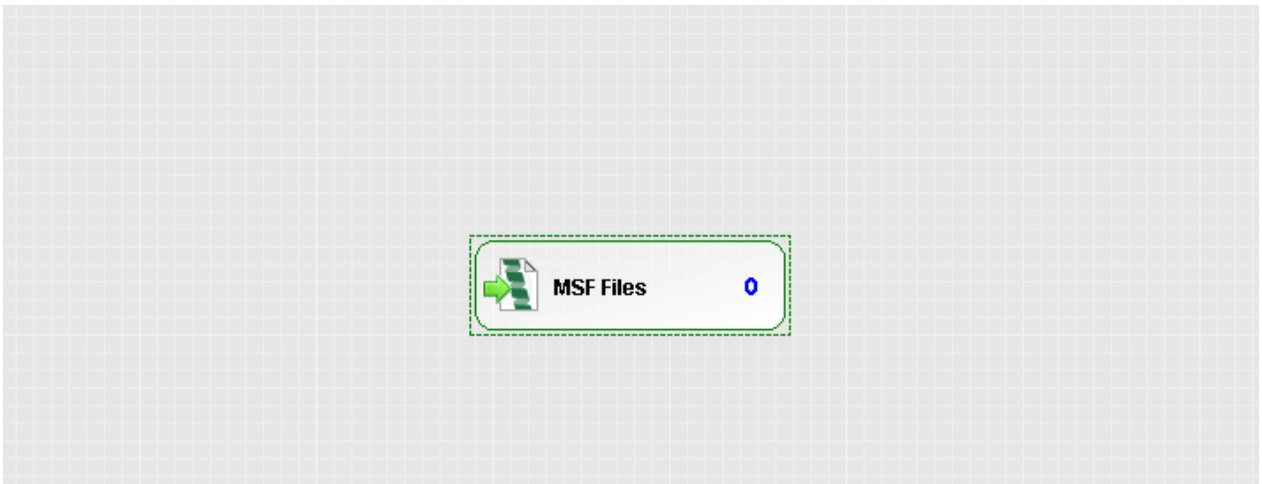
- Restart the Thermo Scientific Proteome Discoverer, navigate to the licencing page and click on Scan for missing features. Subsequently, restart the program once more.

Example Workflow:

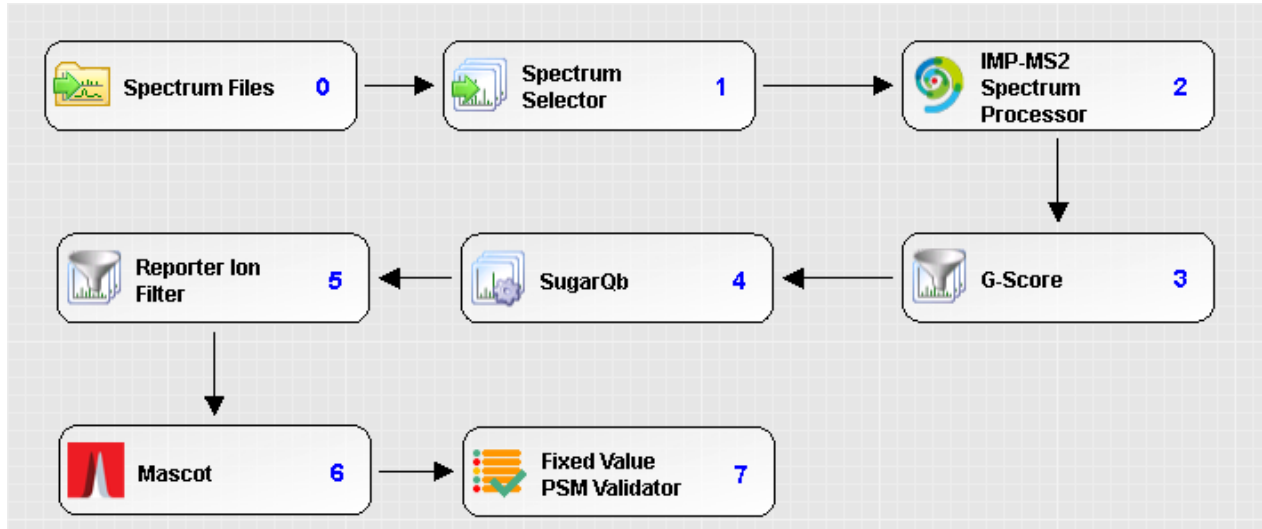
In Proteome Discoverer there are two different type of Workflows the Consensus Workflow and the Processing Workflow.

- Download the test data file “LUMOS_SugarQb_Test_humanPlasma_HCDonly.raw” from: www.imba.oeaw.ac.at/SugarQb . This data has been generated by analyzing IP-HILIC-enriched, tryptic glycopeptides derived from a chemically de-sialylated human plasma, using HCD on an OrbiTrap Fusion LUMOS instrument.
- After Installation of the SugarQb Nodes, in Thermo Scientific Proteome Discoverer 2.1., create the following Workflow. Parameter settings of the respective Nodes are detailed below.

Consensus Workflow:



Processing Workflow:



Recommended Settings & Parameters:

Spectrum Selector:

N.B.: The default settings of the Spectrum Selector Node were modified, to also allow “higher” mass precursor ions(i.e. up to 10.000 Da) to be analyzed.

Parameters	
Show Advanced Parameters	
▾ 1. General Settings	
Precursor Selection	Use MS1 Precursor
▾ 2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
Ignore Specified Scans	
Lowest Charge State	0
Highest Charge State	0
Min. Precursor Mass	350 Da
Max. Precursor Mass	10000 Da
Total Intensity Threshold	0
Minimum Peak Count	1
▾ 3. Scan Event Filters	
Mass Analyzer	(Not specified)
MS Order	Is MS2
Activation Type	(Not specified)
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Is Full
Polarity Mode	Any
▾ 4. Peak Filters	
S/N Threshold (FT-only)	1.5
▾ 5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacer	Automatic
Unrecognized Mass Analyzer Replac	ITMS
Unrecognized MS Order Replac	MS2
Unrecognized Activation Type Replac	CID
Unrecognized Polarity Replacer	+
Unrecognized MS Resolution@	60000
Unrecognized MSn Resolution@	30000

MS2 – Spectrum Processor

This Node provides two MS2-spectrum preprocessing steps: Deisotoping of isotopic clusters and charge-deconvolution. For this, spectra are searched for isotopic clusters by determining the distances in m/z values between pairs of peaks. For every cluster detected, only the monoisotopic peaks remain in the spectrum, other peaks are removed.

Subsequently, the spectra are deconvolved to charge state 1. Every peak with a charge state greater than 1 will be removed from the spectrum and replaced by a peak at the corresponding singly-charged m/z -position with the same intensity. Note that the algorithm only works on peaks having charge state information available. For a more detailed description of the algorithm, please refer to:

<http://ms.imp.ac.at/?goto=pd-nodes>

In this exemplary workflow, the following parameter settings are recommended:

Parameters	
<input type="button" value="Hide Advanced Parameters"/>	
1. General Settings	
Perform De-Isotoping	True
Select DeIsotoping Method	Standard
Isotope Distance Deviation Tolerance	25 mmu
Minimal Isotope Ratio	0.3
Use Adaptive Isotope Distance Deviation Toler	True
Deisotope Reporter Region	True
Perform Charge De-Convolution	True
Select Charge-Deconvolution method	Standard
2. Averagine Modelling Settings	
Modelling Tolerance	0.5
Use Relative Intensity Threshold	False
Intensity Threshold	0
Apply Adaptive Modelling	False
Use Pattern Scoring (Best - Fit Isotope Pattern)	False
3. MS1 Preprocessing Settings	
Recalculate Precursor mass from MS1	False
Use 3d Peaks	True
3d peak-picking tolerance	5 ppm
Minimum profile points for 2d peak	5
Detect 3d split-peak	True
Regression window	4
Number of Skip-Scans	1
Use Isotopes	True
Isotope Distance Tolerance	5 mmu
Use Averagine Modeling	True

G-Score (optional):

The G-Score Node filters MS2 spectra based on the occurrence and intensity of various glycan-derived oxonium ions (for more details see Stadlmann J., Taubenschmid J., et al. Nature (2017)), and thus allows for a more efficient analysis of glycopeptides. Optimal threshold settings need to be empirically established for each instrument acquisition method. In this example, a G-Score threshold of 0.4 was used. N.B. the use of this Node is optional.

Parameters	
<input type="button" value="Hide Advanced Parameters"/>	
▲ 1. Scoring Parameters	
Mass Tolerance	5 ppm
G-Score Threshold	0.4
Filter G-Scores >= Threshold	True

SugarQb:

The SugarQb Node focuses on the identification of the potential [peptide + HexNAc] + - fragment ions within MS/MS spectra. For this, the precursor-ion masses of a given MS/MS spectrum are iteratively reduced by all masses present in a glycan-composition database, minus the mass of one HexNAc residue (i.e. 203.0794 amu). This approach generates a set of theoretical [peptide + HexNAc] + -fragment ion masses, which are then tried to be matched within the MS/MS spectrum. In cases where an experimental peak matches a theoretical [peptide + HexNAc] + -fragment, the concomitant presence of the corresponding potential [peptide]+ -fragment ion is verified. Only if both peaks are detected, the given spectrum is duplicated with its precursor-ion mass set to the mass of the respective potential [peptide + HexNAc] + fragment-ion (for more details see Stadlmann J., Taubenschmid J., et al. Nature (2017)).

Note, that in this exemplary workflow, charge-deconvoluted MS2 spectra (i.e. all fragment ions are expected to be of charge state 1) are analyzed and thus only charge state 1 is allowed. The .txt Glyco Database File used in this example can be downloaded at:

www.imba.oeaw.ac.at/SugarQb

The following SugarQb parameter settings are recommended:

Parameters	
Show Advanced Parameters	
1. Processing Criteria	
Mass Tolerance	20 mmu
Intensity Threshold	0
Top N Peaks	0
Allowed Charge States	1
Glyco Database File Selection	C:\Users\venes.sakalli\Desktop\Enes\ThermoGlyc\NO-GlycanDB_final+17Da_
Enforce Peptide Peak Match	True
Enforce Peptide + 2 * HexNAc Peak Match	False

Mass Tolerance
The mass tolerance of the reporter ions.

Reporter Ion-Filter(Optional):

This Node enables the filtering/removal of usually highly abundant, glycan-related fragment ions from MS2 spectra. Reporter Ion masses to be completely removed from the MS2 dataset can also be defined in a separate .txt file (i.e. Reporter Ion File Selection). The Reporter Ion File used in this example can be downloaded at: www.imba.oeaw.ac.at/SugarQb. N.B. the use of this Node is optional.

Parameters	
Show Advanced Parameters	
1. Filter Criteria	
Reporter Ion(s) Mass	204.08667
Top N Peaks	0
Mass Tolerance	5 ppm
Intensity Threshold	0.1
Reporter Ion File Selection	C:\Users\enes.sakalli\Desktop\Enes\ThermoGlyc

Reporter Ion(s) Mass

Determines which reporter ion masses have to be considered for filtering.

MS/MS Search Engine Settings & Parameters:

For the eventual identification of the glycopeptide amino-acid sequences, all MS2 spectra generated by the SugarQb Node (i.e. those with the original and those with the modified precursor-ion masses) are searched against a concatenated forward and decoy database of the Uniprot human reference proteome set, considering HexNAc (and its neutral loss of 203.079373 amu) as a variable modification to any asparagine, serine and threonine residue.

Here, the use of MASCOT and SEQUEST-HT, are exemplified. Of Note, an in-house developed MS/MS search engine, MS Amanda, is freely available at: <http://ms.imp.ac.at/?goto=pd-nodes> Irrespective of the MS/MS search engine employed, the resulting peptide-spectrum matches (PSMs) are then manually filtered. For this, only the bestscoring PSMs of each spectrum group (i.e. comprising the MS/MS spectrum with the original precursor-ion mass and all its duplicates with the respectively modified precursor ion masses) are kept and filtered to an estimated false discovery rate (FDR) of 1%, employing the standard “target-decoy approach” (Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4, 207-214 (2007)).

Importantly, PSM consolidation prior to manual FDR-filtering can be performed using the “Glyco-Filter” Node (see below).

MASCOT:

Recommended MASCOT search parameter settings are listed below. Of note, MASCOT provides additional options to optimize the search engines performance in the identification of glycopeptide amino-acid sequences (e.g. handling of the dominant neutral loss of the glycan portion upon HCD fragmentation or scoring only singly charged fragment ions). Examples of such adjustments are described in the Annex section.

1. Input Data	
Protein Database	human_uniprot_comb
Enzyme Name	Trypsin
Maximum Missed Cleavage Sites	2
Instrument	ESI QUAD 1+
Taxonomy	All entries
2. Tolerances	
Precursor Mass Tolerance	20 mmu
Fragment Mass Tolerance	20 mmu
Use Average Precursor Mass	False
3. Modification Groups	
From Quan Method	
4. Dynamic Modifications	
Show All Modifications	False
1. Dynamic Modification	HexNAc(NL) (NST)
2. Dynamic Modification	Oxidation (C)
3. Dynamic Modification	
4. Dynamic Modification	
5. Dynamic Modification	
6. Dynamic Modification	
7. Dynamic Modification	
8. Dynamic Modification	
9. Dynamic Modification	
5. Static Modifications	
1. Static Modification	Carbamidomethyl (C)
2. Static Modification	
3. Static Modification	
4. Static Modification	
5. Static Modification	
6. Static Modification	

SEQUEST-HT:

Recommended SEQUEST-HT search parameter settings are listed below.

1. Input Data	
Protein Database	
Enzyme Name	Trypsin (Full)
Max. Missed Cleavage Sites	2
Min. Peptide Length	6
Max. Peptide Length	144
2. Tolerances	
Precursor Mass Tolerance	20 ppm
Fragment Mass Tolerance	0.025 Da
Use Average Precursor Mass	False
Use Average Fragment Mass	False
3. Spectrum Matching	
Use Neutral Loss a Ions	True
Use Neutral Loss b Ions	True
Use Neutral Loss y Ions	True
Use Flanking Ions	True
Weight of a Ions	0
Weight of b Ions	1
Weight of c Ions	0
Weight of x Ions	0
Weight of y Ions	1
Weight of z Ions	0
4. Dynamic Modifications	
Max. Equal Modifications Per Peptide	3
1. Dynamic Modification	HexNAc / +203.079 Da (N, S, T)
2. Dynamic Modification	Oxidation / +15.995 Da (M)
3. Dynamic Modification	None
4. Dynamic Modification	None
5. Dynamic Modification	None
6. Dynamic Modification	None
5. Dynamic Modifications (peptide terminus)	
1. N-Terminal Modification	None
2. N-Terminal Modification	None
3. N-Terminal Modification	None
1. C-Terminal Modification	None
2. C-Terminal Modification	None
3. C-Terminal Modification	None
6. Dynamic Modifications (protein terminus)	
1. N-Terminal Modification	None
2. N-Terminal Modification	None
3. N-Terminal Modification	None
1. C-Terminal Modification	None
2. C-Terminal Modification	None
3. C-Terminal Modification	None
7. Static Modifications	
Peptide N-Terminus	None
Peptide C-Terminus	None
1. Static Modification	Carbamidomethyl / +57.021 Da (C)
2. Static Modification	None
3. Static Modification	None
4. Static Modification	None
5. Static Modification	None
6. Static Modification	None

MS-AMANDA:

Recommended SEQUEST-HT search parameter settings are listed below.

Parameters	
Show Advanced Parameters	
1. Input Data	
Protein Database	
Enzyme Name	Trypsin (Full)
Missed Cleavages	2
MS1 tolerance	5 ppm
MS2 tolerance	0.02 Da
2. Static Modifications	
1. Static Modification	Carbamidomethyl / +57.021 Da (C)
2. Static Modification	None
3. Static Modification	None
Static Peptide N-Terminal Modification	None
Static Peptide C-Terminal Modification	None
Static Protein N-Terminal Modification	None
3. Dynamic Modifications	
1. Dynamic Modification	Oxidation / +15.995 Da (M)
2. Dynamic Modification	HexNAc / +203.079 Da (N, S, T)
3. Dynamic Modification	None
21. Dynamic Peptide N-Terminal Modification	None
22. Dynamic Peptide C-Terminal Modification	None
23. Dynamic Protein N-Terminal Modification	None
Protein Database	
The sequence database to be searched.	

ptmRS(Optional)

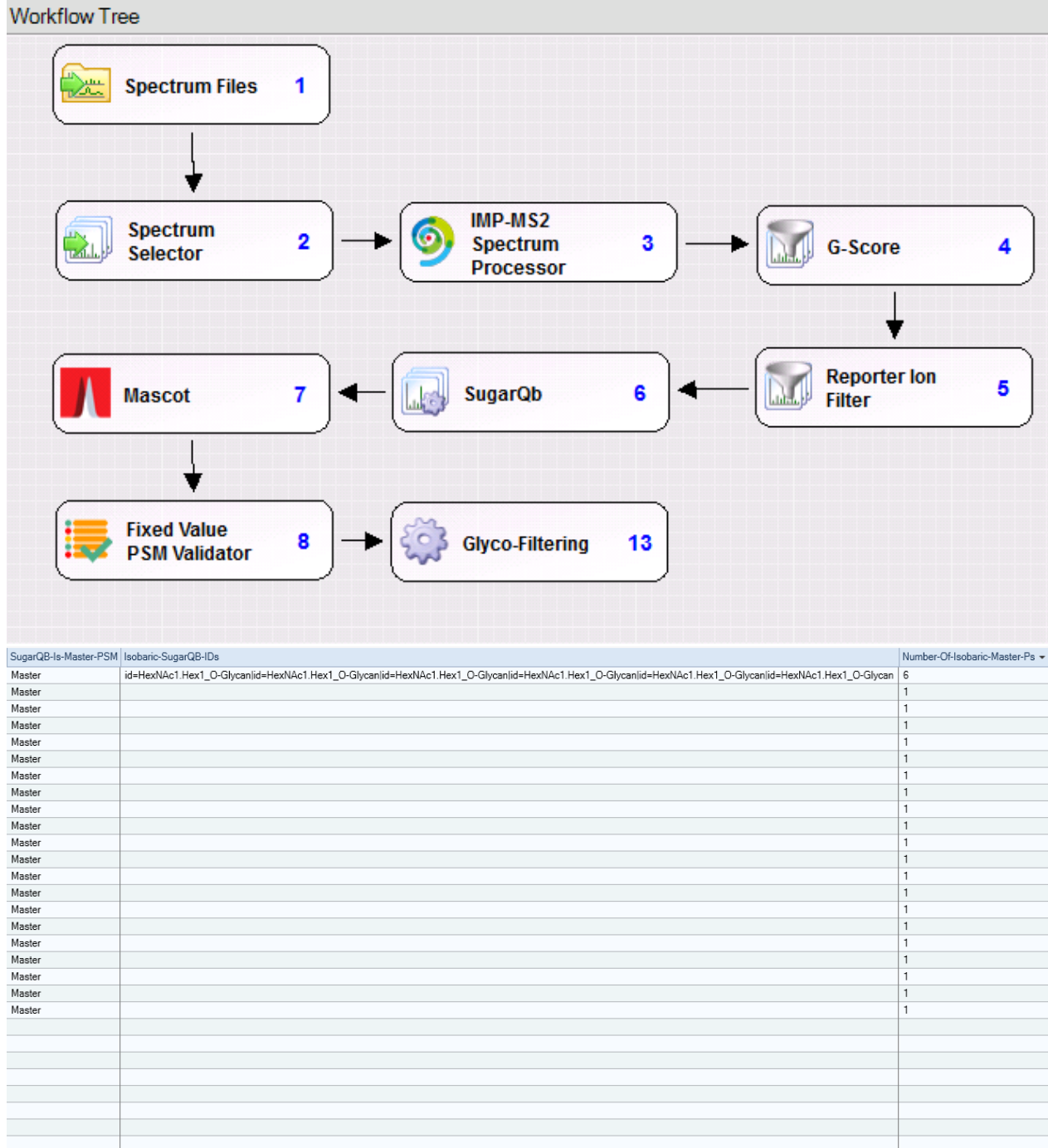
Generally, this tool enables automated and confident localization of modification sites within validated peptide sequences. It calculates individual probability values for each putatively modified site based on the given MS/MS data. ptmRS can also be used to localize N- glycosylation sites. For further information on the algorithm of the software, please refer to Taus T., et al. (2011) *Universal and Confident Phosphorylation Site Localization Using phosphoRS*.

Parameters	
Show Advanced Parameters	
▲ 1. Scoring	
PhosphoRS Mode	False
Use Diagnostic Ions	True

PhosphoRS Mode
If this parameter is set to 'true' then ptmRS will localize only phosphorylation sites while the positions of all other PTMs are based on the search engine's identification. 'False' would indicate that all variable modifications will be localized in parallel.

Additional Workflow with the Glyco-Filter:

The additional Node Glyco-Filter is for annotating the best Psm as Master PSM for Glyco-Data, as well when there are several other Psm which are having the same Ion-Score, it will summarize all the Glycan-IDs of all Psms with the same Ion-Score.



Anticipated Results using the sample data file provided:

The Thermo Scientific Proteome Discoverer 2.1 result files (.msf), the exported Excel workbook, and a manually filtered result-file (.csv) can be downloaded from:

www.imba.oeaw.ac.at/SugarQb .

Contact: SugarQb@imba.oeaw.ac.at

Annex

Alternative MASCOT Parameters (optional)

Instruments Settings:

Instruments		
Ion series	Default	ESI QUAD 1+
1+	X	X
2+	X	
2+ (precursor>3+)		
immonium		
a	X	
a*	X	
a0		
b	X	X
b*	X	X
b0		X
c		
x		
y	X	X
y*	X	X
y0		X
z		
yb		
ya		
y must be significant		
y must be highest score		
z+1		
d		
v		
w		
z+2		
Minimum mass		
Max mass	700	
		Delete
	Edit	Edit

Modification Settings "HexNAc(NL)":

Edit Modification :HexNAc(NL)

Name	
Title	HexNAc(NL)
Fullname	N-Acetylhexosamine Asparagine vs Serine/Threonine
Delta Specificity Ignore Masses Misc References	
Delta	
Monoisotopic	203.079373
Average	203.1925
Composition	HexNAc
Symbols	13C <input type="button" value="v"/> 1 <input type="button" value="v"/> <input type="button" value="Add"/>

Edit Modification :HexNAc(NL)

Name	
Title	HexNAc(NL)
Fullname	N-Acetylhexosamine Asparagine vs Serine/Threonine
Delta Specificity Ignore Masses Misc References	
Specificity	
Specificity	Site <input type="text" value="T"/> Position <input type="text" value="Anywhere"/> Copy Delete Hide Details
Classification	<input type="text" value="Other glycosylation"/> Hidden <input type="checkbox"/> Group <input type="text" value="1"/>
Notes	<input type="text"/>
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Monoisotopic: 203.079373 Average: 203.1925	
Composition	<input type="text" value="HexNAc"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Composition	<input type="text"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
New Neutral Loss	
<hr/>	
Specificity	Site <input type="text" value="S"/> Position <input type="text" value="Anywhere"/> Copy Delete Hide Details
Classification	<input type="text" value="Other glycosylation"/> Hidden <input type="checkbox"/> Group <input type="text" value="1"/>
Notes	<input type="text"/>
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Monoisotopic: 203.079373 Average: 203.1925	
Composition	<input type="text" value="HexNAc"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Composition	<input type="text"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
New Neutral Loss	
<hr/>	
Specificity	Site <input type="text" value="N"/> Position <input type="text" value="Anywhere"/> Copy Delete Hide Details
Classification	<input type="text" value="Other glycosylation"/> Hidden <input type="checkbox"/> Group <input type="text" value="1"/>
Notes	<input type="text"/>
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Monoisotopic: 203.079373 Average: 203.1925	
Composition	<input type="text" value="HexNAc"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
Neutral loss	<input checked="" type="radio"/> Scoring <input type="radio"/> Satellite <input type="radio"/> Peptide <input type="radio"/> Required Peptide Delete
Composition	<input type="text"/> Symbols <input type="text" value="13C"/> <input type="text" value="1"/> Add
New Neutral Loss	

Delta	Specificity	Ignore Masses	Misc	References
Ignore Masses				
Ignore Mass 1	Monoisotopic 204.086649 Average 204.1999		Delete	
Composition	<input type="text" value="H HexNAc e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 2	Monoisotopic 186.076084 Average 186.1846		Delete	
Composition	<input type="text" value="H HexNAc Water(-1) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 3	Monoisotopic 168.065519 Average 168.1694		Delete	
Composition	<input type="text" value="H HexNAc Water(-2) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 4	Monoisotopic 144.065519 Average 144.1480		Delete	
Composition	<input type="text" value="C(6) H(10) N O(3) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 5	Monoisotopic 138.054955 Average 138.1434		Delete	
Composition	<input type="text" value="C(7) H(8) N O(2) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 6	Monoisotopic 126.054955 Average 126.1327		Delete	
Composition	<input type="text" value="C(6) H(8) N O(2) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 7	Monoisotopic 163.060100 Average 163.1480		Delete	
Composition	<input type="text" value="H Hex e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 8	Monoisotopic 145.049535 Average 145.1327		Delete	
Composition	<input type="text" value="H Hex Water(-1) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 9	Monoisotopic 127.038970 Average 127.1174		Delete	
Composition	<input type="text" value="H Hex Water(-2) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 10	Monoisotopic 366.139472 Average 366.3405		Delete	
Composition	<input type="text" value="H Hex HexNAc e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 11	Monoisotopic 528.192296 Average 528.4811		Delete	
Composition	<input type="text" value="H Hex(2) HexNAc e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 12	Monoisotopic 292.102693 Average 292.2620		Delete	
Composition	<input type="text" value="C(11) H(18) N O(8) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 13	Monoisotopic 274.092128 Average 274.2467		Delete	
Composition	<input type="text" value="C(11) H(18) N O(8) Water(-1) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 14	Monoisotopic 657.235438 Average 657.5956		Delete	
Composition	<input type="text" value="C(11) H(18) Hex HexNAc N O(8)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 15	Monoisotopic 243.026430 Average 243.1279		Delete	
Composition	<input type="text" value="H(2) Hex O(3) P e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add
Ignore Mass 16	Monoisotopic 225.015866 Average 225.1126		Delete	
Composition	<input type="text" value="H(2) Hex O(3) P Water(-1) e(-1)"/>			
	Symbols	<input type="text" value="13C"/>	<input type="text" value="1"/>	Add

