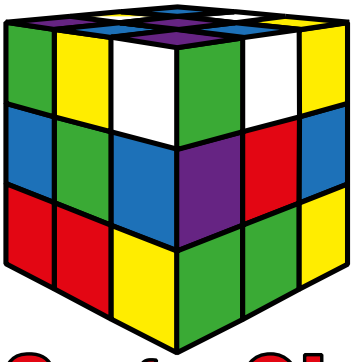


HOWTO, example workflow and data files.

(Version 20-09-2017)

Introduction:

SugarQb is a collection of software tools (Nodes) which enable the automated identification of intact glycopeptides from HCD-MS/MS data sets, using common MS/MS search engines (e.g. MASCOT, SEQUEST-HT) in the Proteome Discoverer 1.4 environment.

SugarQb is freely available to all researchers.

For further information on the algorithm, please refer to the corresponding publication Stadlmann J., Taubenschmid J., 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 who to download, install and test SugarQb, analyzing an 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@imba.oeaw.ac.at

Download and Installation:

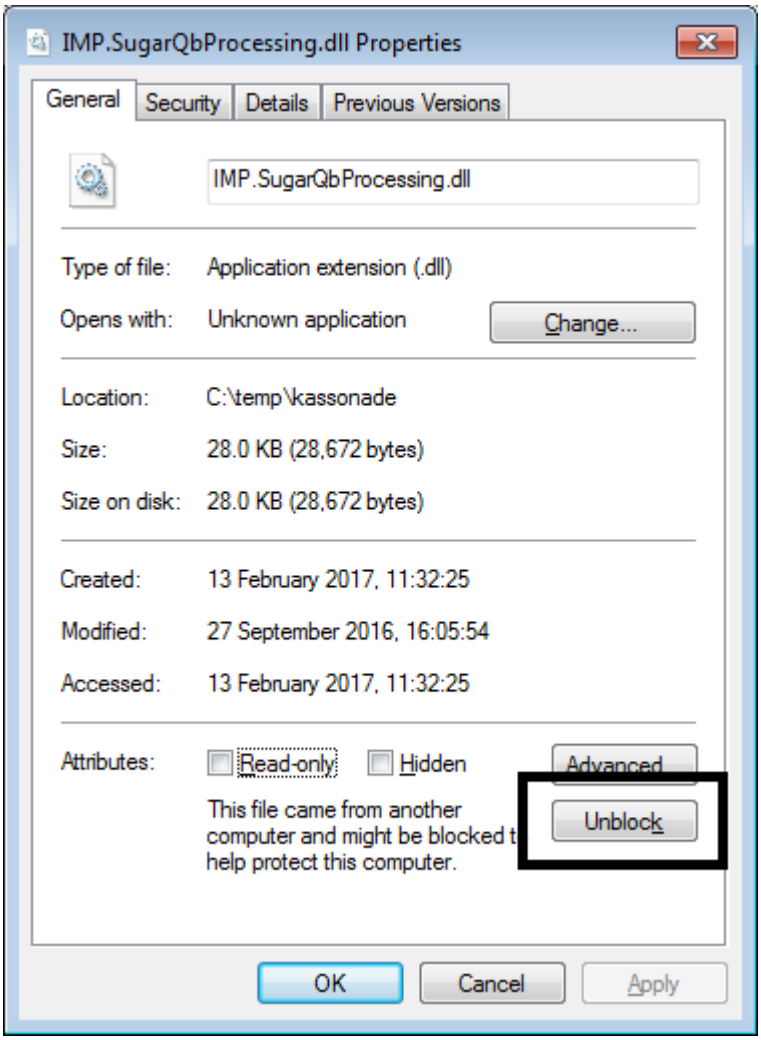
- Download SugarQb for Thermo Scientific Proteome Discoverer 1.4. using the following URL: www.imba.oeaw.ac.at/SugarQb

- 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 folder 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.

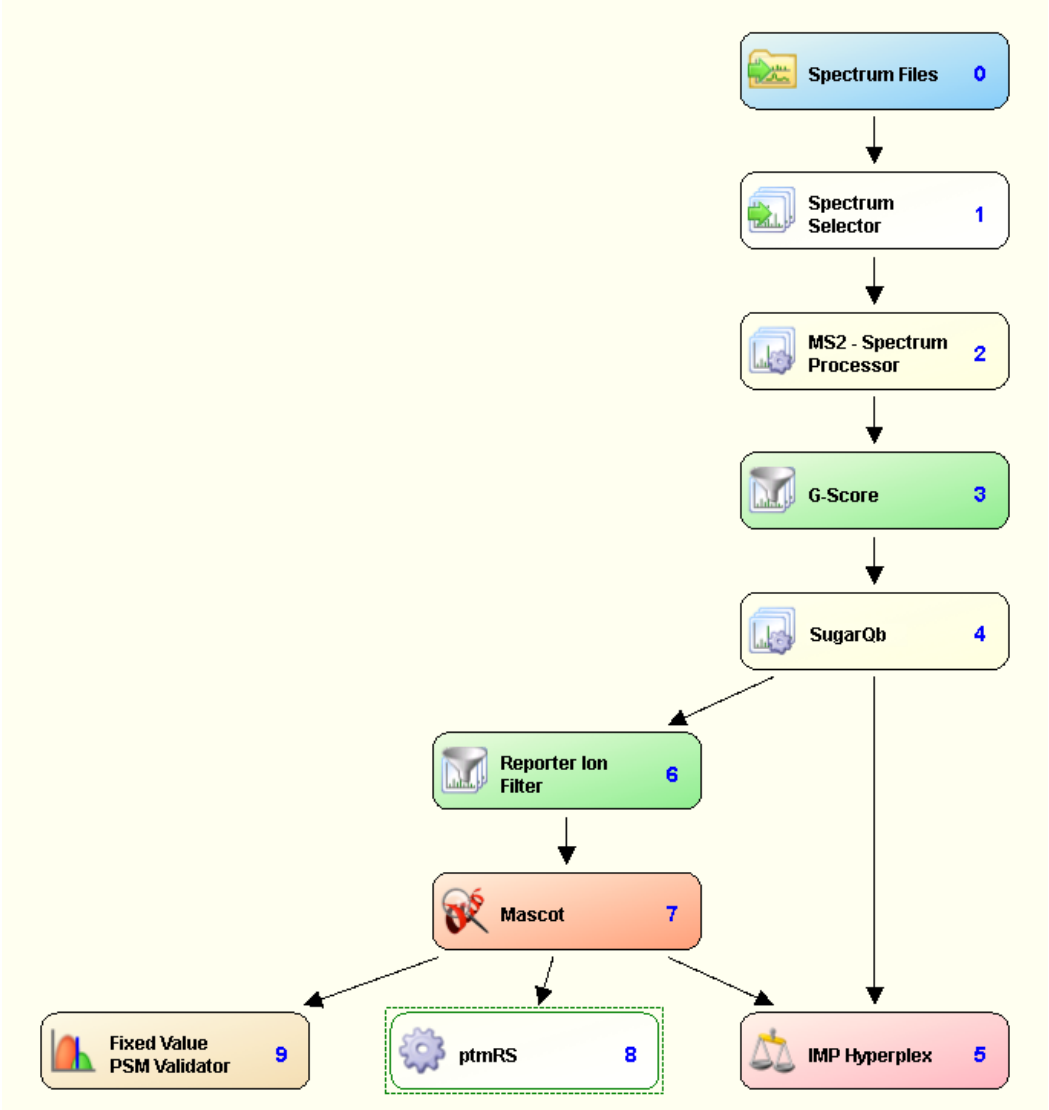


- Restart Thermo Scientific Proteome Discoverer, navigate to the licensing page and click on Scan for Missing Features. Subsequently, restart the program once more.

Example 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 a OrbiTrap Fusion LUMOS instrument.

- After Installation of the SugarQb Nodes, in Thermo Scientific Proteome Discoverer 1.4., create the following Workflow. Parameter settings of the respective Nodes are detailed below.



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

Hide Advanced Parameters

1. General Settings	
Precursor Selection	Use MS1 Precursor
Use New Precursor Reevaluation	True
2. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	0
First Scan	0
Last Scan	0
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	Any
MS Order	Is MS2
Activation Type	Any
Min. Collision Energy	0
Max. Collision Energy	1000
Scan Type	Is Full
Ionization Source	Any
Polarity Mode	Any
4. Peak Filters	
S/N Threshold (FT-only)	1.5
5. Replacements for Unrecognized Properties	
Unrecognized Charge Replacements	Automatic
Unrecognized Mass Analyzer Replacements	ITMS
Unrecognized MS Order Replacements	MS2
Unrecognized Activation Type Replacements	CID
Unrecognized Polarity Replacements	+
6. Just for Testing	
Precursor Clipping Range Before	2.5 Da
Precursor Clipping Range After	5.5 Da

1. General Settings

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

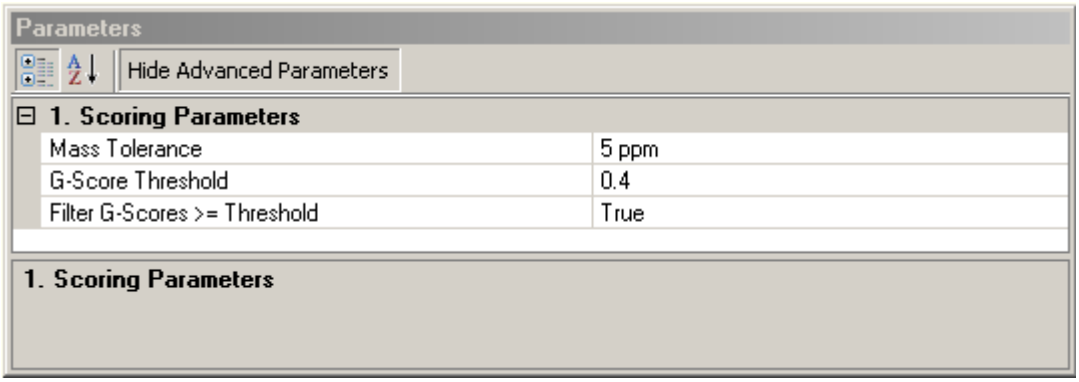
Hide Advanced Parameters

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 Tolerance	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 Search)	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

1. General Settings

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.

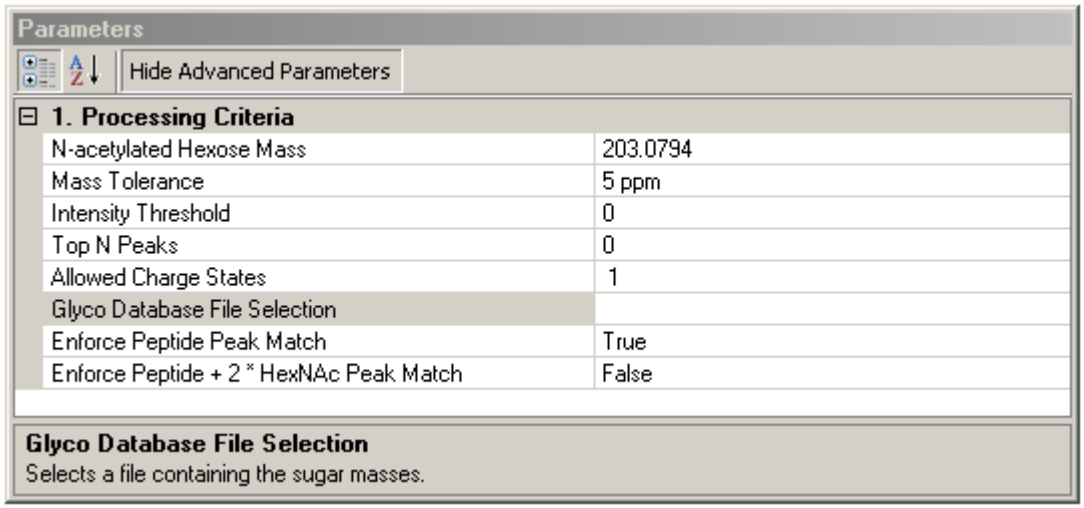


SugarQb:

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)).

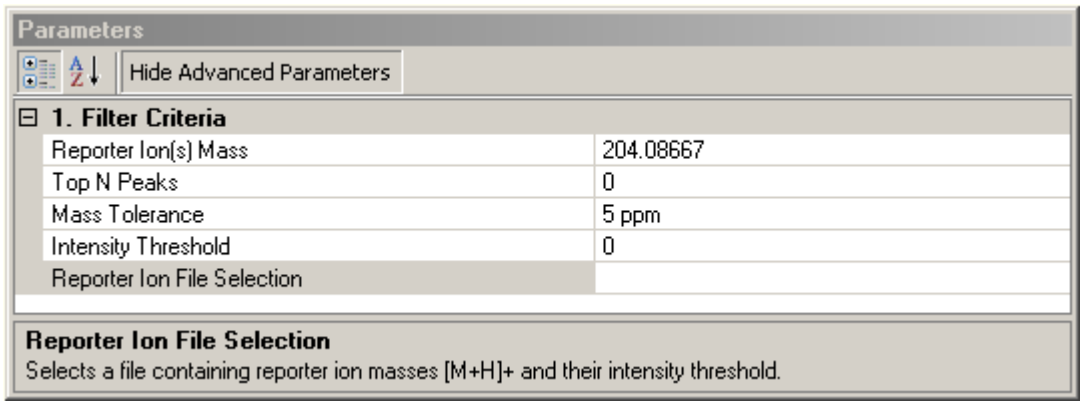
Note, that in this exemplary workflow, charge-deconvoluted MS2 spectra (i.e. all fragment ions are expected to be of charge state 1) are analyzed and thus only charge state 1 is allowed. The .txt Glyco Database File used in this example can be downloaded at: www.imba.oeaw.ac.at/SugarQb

The following SugarQb parameter settings are recommended:



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.



MS/MS Search Engine Settings & Parameters:

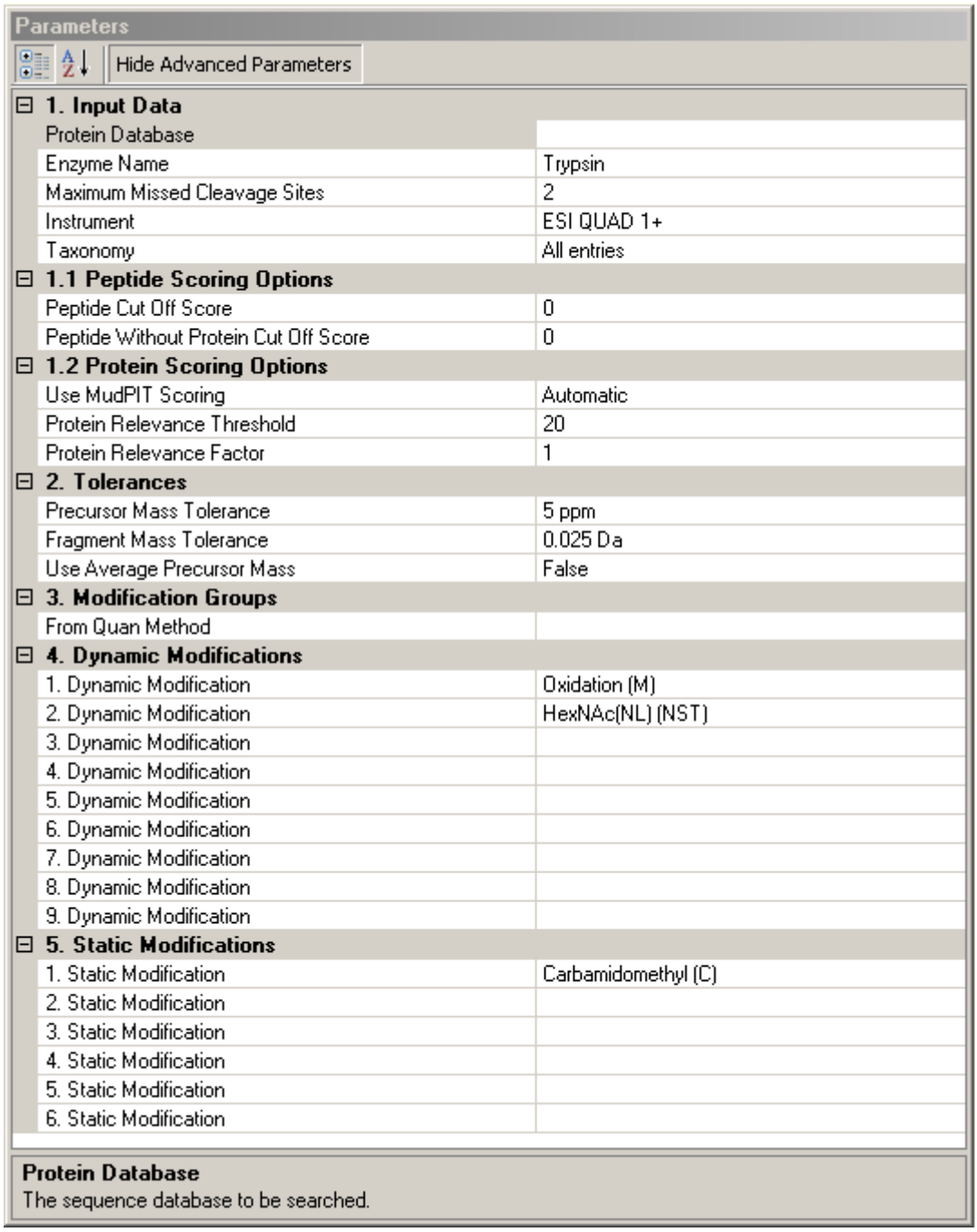
For the eventual identification of the glycopeptide amino-acid sequences, all MS2 spectra generated by the SugarQb Node (i.e. those with the original and those with the modified precursor-ion masses) are searched against a concatenated forward and decoy database of the Uniprot human reference proteome 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 best scoring 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)).

Currently, PSM filtering is performed, after exporting the search results from Thermo Scientific Proteome Discoverer 1.4. to a .csv file, using a Perl script. These scripts can be downloaded at: www.imba.oeaw.ac.at/SugarQb. N.B. the use of the FDR-filtering script is optional.

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.



Parameters

Hide Advanced Parameters

1. Input Data	
Protein Database	
Enzyme Name	Trypsin
Maximum Missed Cleavage Sites	2
Instrument	ESI QUAD 1+
Taxonomy	All entries
1.1 Peptide Scoring Options	
Peptide Cut Off Score	0
Peptide Without Protein Cut Off Score	0
1.2 Protein Scoring Options	
Use MudPIT Scoring	Automatic
Protein Relevance Threshold	20
Protein Relevance Factor	1
2. Tolerances	
Precursor Mass Tolerance	5 ppm
Fragment Mass Tolerance	0.025 Da
Use Average Precursor Mass	False
3. Modification Groups	
From Quan Method	
4. Dynamic Modifications	
1. Dynamic Modification	Oxidation (M)
2. Dynamic Modification	HexNAc(NL) (NST)
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	

Protein Database
The sequence database to be searched.

SEQUEST-HT:

Recommended SEQUEST-HT search parameter settings are listed below.

Parameters

Hide Advanced Parameters

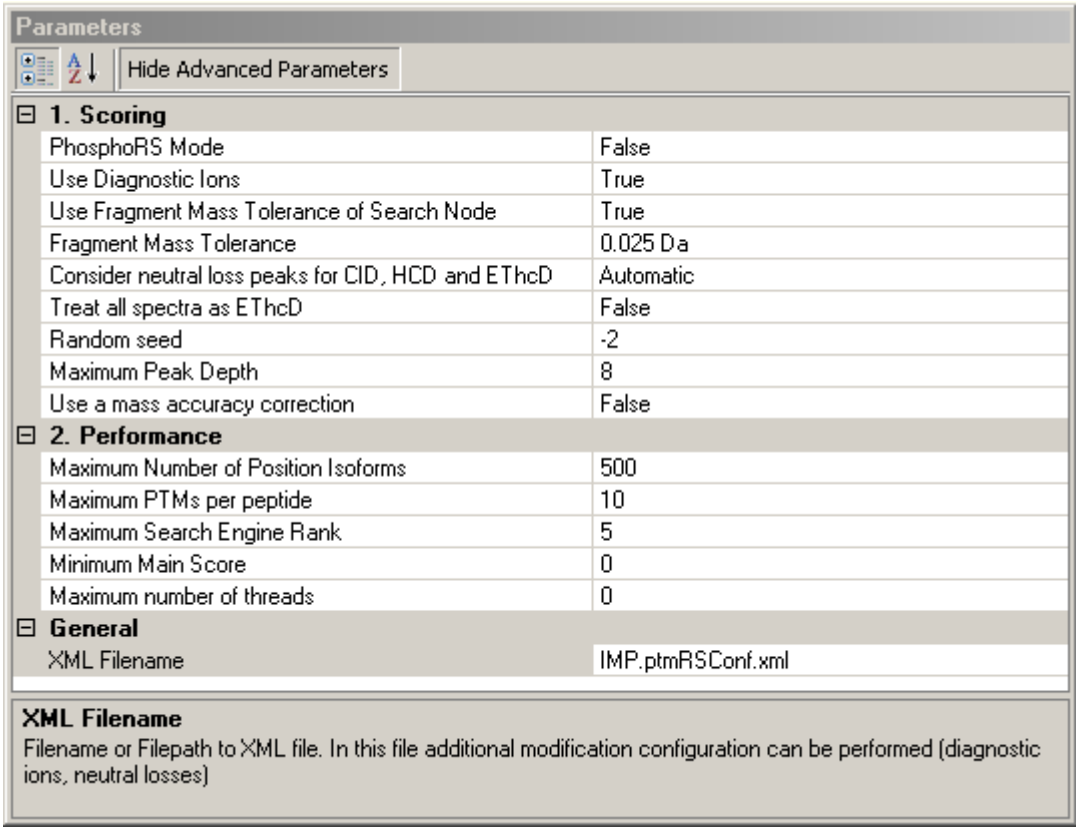
1. Input Data	
Protein Database	
Enzyme Name	Trypsin (Full)
Max. Missed Cleavage Sites	2
Min. Peptide Length	6
Max. Peptide Length	144
2. Scoring Options	
Max. Delta Cn	0.05
Max. Number of Peptides Reported	5
3. Tolerances	
Precursor Mass Tolerance	5 ppm
Fragment Mass Tolerance	0.025 Da
Use Average Precursor Mass	False
Use Average Fragment Mass	False
4. 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
5. Dynamic Modifications	
Max. Equal Modifications Per Peptide	3
Max. Dynamic Modifications Per Peptide	4
N-Terminal Modification	None
C-Terminal Modification	None
1. Dynamic Modification	Oxidation / +15.995 Da (M)
2. Dynamic Modification	HexNAc / +203.079 Da (N, S, T)
3. Dynamic Modification	None
4. Dynamic Modification	None
5. Dynamic Modification	None
6. Dynamic Modification	None
6. 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

Protein Database
The sequence database to be searched.

ptmRS (Optional):

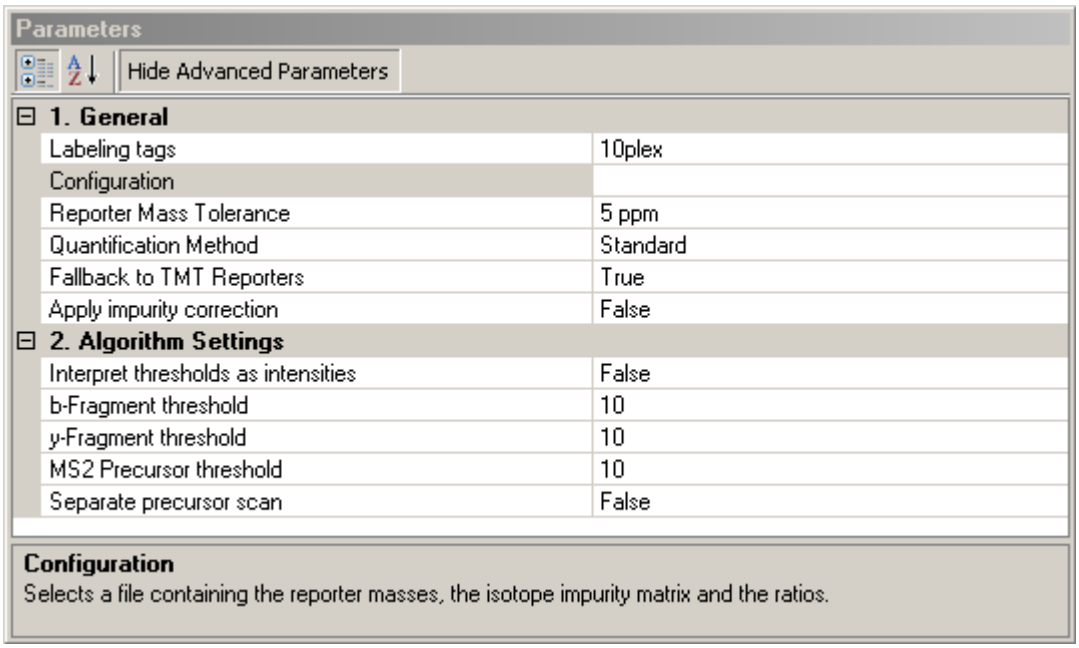
Generally, this tool enables automated and confident localization of modification sites within validated peptide sequences. It calculates individual probability values for each putatively modified site based on the given MS/MS data. ptmRS can also be used to localize N-glycosylation sites. For further information on the algorithm of the software, please refer to: http://ms.imp.ac.at/?goto=pd-nodes

The .xml configuration file used in this exemplary workflow can be downloaded at: www.imba.oeaw.ac.at/SugarQb. N.B. the use of this Node is optional.



IMP Hyperplex (optional):

The IMP Hyperplex has far-reaching peptide-quantification capabilities. Importantly, this node allows to also extract quantitative data of individual, user-defined m/z-bins. This information can e.g. be used to manually inspect the agreement between glycan composition and the presence of diagnostic oxonium ions. Ion masses to be considered are defined in a separate .txt configuration file. The configuration file used in this example can be downloaded at: www.imba.oeaw.ac.at/SugarQb. N.B. the use of this Node is optional.



Anticipated Results using the sample data file provided:

The Thermo Scientific Proteome Discoverer 1.4. 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	
	Edit	Delete 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 1 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: T — Position: Anywhere — Copy | Delete | Hide Details

Classification: Other glycosylation — Hidden ☐ — Group: 1

Notes:

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Monoisotopic: **203.079373** Average: **203.1925**

Composition: HexNAc — Symbols: 13C 1 Add

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Composition: — Symbols: 13C 1 Add

New Neutral Loss

Specificity — Site: S — Position: Anywhere — Copy | Delete | Hide Details

Classification: Other glycosylation — Hidden ☐ — Group: 1

Notes:

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Monoisotopic: **203.079373** Average: **203.1925**

Composition: HexNAc — Symbols: 13C 1 Add

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Composition: — Symbols: 13C 1 Add

New Neutral Loss

Specificity — Site: N — Position: Anywhere — Copy | Delete | Hide Details

Classification: Other glycosylation — Hidden ☐ — Group: 1

Notes:

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Monoisotopic: **203.079373** Average: **203.1925**

Composition: HexNAc — Symbols: 13C 1 Add

Neutral loss: ◉ Scoring ○ Satellite ○ Peptide ○ Required Peptide — Delete

Composition: — Symbols: 13C 1 Add

New Neutral Loss

Delta | Specificity | Ignore Masses | Misc | References

Ignore Masses

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