

pMatchGlyco User Guide

(Last updated: 7/16/2018)

Outline

1. Overview	1
2. Installation.....	1
3. Quick Start.....	1
3.1 Preparation to use pMatchGlyco.....	1
3.2 Running pMatchGlyco.	3
4. Output.....	6
5. FAQ.....	8

1. Overview

pMatchGlyco is a new open spectral library search tool for intact N-linked glycopeptide identifications using higher-energy collisional dissociation (HCD) MS/MS data. As a modified version of [pMatch](#), pMatchGlyco inherits several advantages, such as library spectral optimization with the peptide sequence information, its powerful scoring system and its target-decoy strategy for quality control. The core of pMatchGlyco is written in standard C++ and the interface is implemented in C# in Windows Systems.

2. Installation

A computer with a 64-bit Windows 7 operating system is required to run pMatchGlyco.

1. Download the zipped file [pmatchglyco-1.0.zip](#).
2. Go to the directory where the file pmatchglyco-1.0.zip and unzip it.

3. Quick Start

3.1 Preparation to use pMatchGlyco

- (a) De-glycopeptide and glycopeptide spectra in “MGF” format. We recommend [MSConvert](#) in [ProteoWizard](#) or [pParse](#) to convert raw files into “MGF” format.
- (b) Identification of de-glycopeptide spectra. pMatchGlyco is capable of identified

PSMs derived from any search engine so long as the data follows this tab-delimited syntax:

specName	charge	peptide	modification
----------	--------	---------	--------------

Each field is tab-separated. One row per PSM. Descriptions of what each field is follows along with an example of what a valid value might look like.

- **specName:** This is the name of one spectrum.

Example: sample_1.1.1.2.dta

- **charge:** the charge state of the PSM. This value must be an integer.

Example: 3

- **peptide:** This is the amino acid sequence of PSM.

Example: GNVNFTVSAER

- **modification:** This is a string that indicates what modifications go on what residues of the peptide. The first residue in a peptide is position 1 and the rest increment from there. The position for N-terminal and C-terminal modifications are 0 and peptide_length+1, respectively. If there is no modification identified, set it to be “null”.

The syntax is: position,modified_AA(modification_name);

Example: 4,N(Deamidated[N]);18,C(Carbamidomethyl[C]);

Here is a sample of what a correctly written input file might look like:

specName	charge	peptide	modification
sample_1.1.1.2.dta	2	GNVNFTSVPEHGR	4,N(Deamidated[N]);
sample_1.2.2.1.dta	1	NEEYNK	5,N(Deamidated[N]);
sample_1.3.3.2.dta	2	MVSHHNLTTGATLINEQWLL	6,N(Deamidated[N]);
sample_1.4.4.1.dta	1	NEEYNK	5,N(Deamidated[N]);

The file exported by pBuild or MaxQuant (msms.txt) are also supported. pBuild can be download [here](#).

Please note that the specName should be the same as the spectrum in ‘mgf’ formatted data set and identification results. For example, if you use MSConvert to convert ‘Raw’ to ‘mgf’, the specName should be like

“sample_1.1.1.2”, not including “.dta” in the suffix. Meanwhile, if pParse is selected, the specName should be like “sample_1.1.1.2.0.dta”, adding “.0” before the suffix “.dta”.

(c) Configuration of file “MassTable.ini”. This file contains the mass weight of amino acids, some common modifications and Molecule, and glycans.

Note 1: Make sure that the modification types, which specified in de-glycopeptide spectra identification step, are all specified in “[Modification]” section.

Note 2: We provide a list of N-linked glycans as default. Its glycan name is the monosaccharide composition like “#(Hex)-#(HexNAc)-#(NeuAc)-#(NeuGc)-#dHex”. Users can also provide their own N-linked glycans instead of them. The syntax is: glycan_name=glycan_mass.

Note3: In the software folder, “MassTable.ini” is provided by pGlyco 1.0 [1] and used as default. “MassTable-OVCAR3.ini” is provided by GPQuest when analyzing OVCAR3 data [2]. If user would like to use “MassTable-OVCAR3.ini”, please change its name to be “MassTable.ini”.

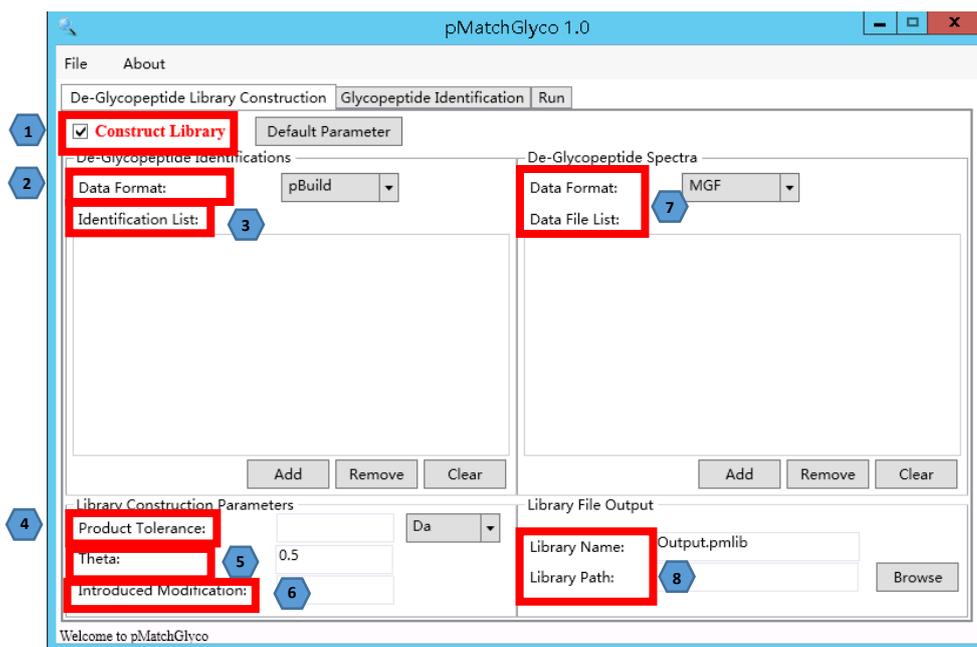
[1] Zeng W F, Liu M Q, Zhang Y, et al. pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD-and CID-MS/MS and MS3[J]. Scientific Reports, 2016, 6: 25102.

[2] Sun S, Shah P, Eshghi S T, et al. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides[J]. Nature biotechnology, 2016, 34(1): 84.

3.2 Running pMatchGlyco.

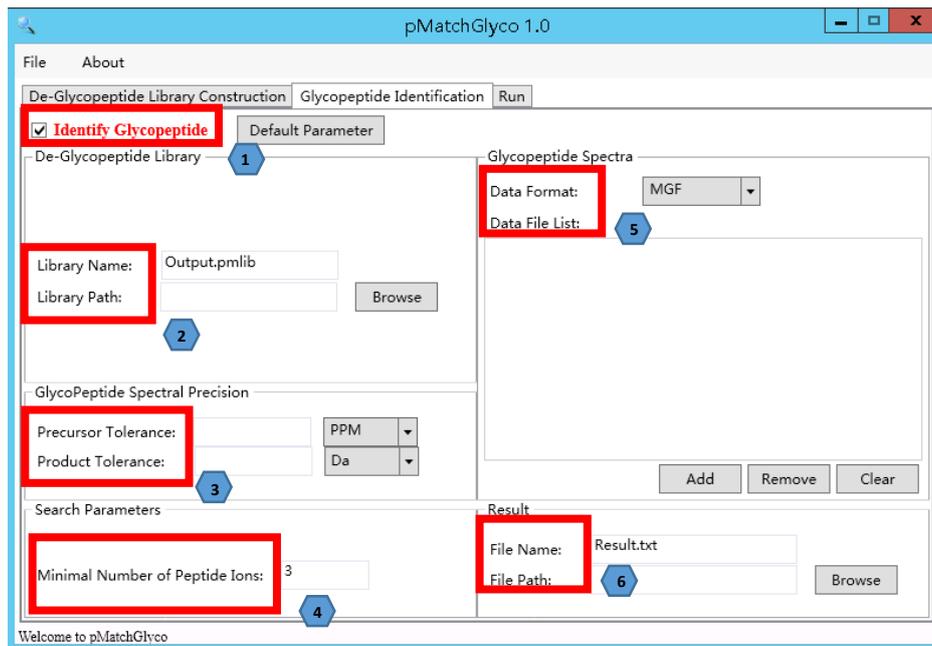
pMatchGlyco uses an interface to specify parameters, and to do the search after clicking the “Run” button. Double click the file  pMatchGlyco 1.0.exe , the interface will appear.

(a) “De-Glycopeptide library Construction” tab page



1. “Construct Library”: Whether to construct library.
2. “Data format” in “De-Glycopeptide identifications” section is the format of previously identified de-glycopeptide spectra. The currently supported formats are: “pBuild”(spectra.txt), “Tab-txt” and “MaxQuant” (msms.txt).
4. “Identification List”: The results files of previously identified spectra.
5. “Product Tolerance” is the m/z tolerance used when annotating product ion peaks in de-glycopeptide spectra.
6. “Theta” is the weight of sequence information used when optimizing library spectra. See our publication for details.
7. “Introduced Modification” is the modification that introduced by PNGase F treatment (Deamidated[N] for pBuild and Deamidation[N] for MaxQuant) or HCD-pd-MS3 technique (HexNAc[N]). Please note that the amino acid ‘N’ should be changed to ‘J’ if you have modified ‘N’ within the N-glycan motif to be ‘J’ in the identification results.
8. “De-Glycopeptide Spectra”: The “Data Format” currently supported “MGF”. The “Data File List” is the list of identified de-glycopeptide spectra.
9. “Library File Output”: the library file name and path, in PMLIB format.

(b) “Glycopeptide Identification” tab page



1. “Identify Glycopeptide”: Whether to identify glycopeptide.
2. “De-Glycopeptide Library” is the file path and name of the reference library, in PMLIB format. (It will be auto filled after “De-Glycopeptide Library Construction” operation)
3. “Glycopeptide Spectral Precision”: “Precursor Tolerance” and “Product Tolerance” are the precursor and product tolerances in glycopeptide spectra, respectively.
4. “Search Parameters”: “Minimal Number of Peptide Ions” is the number threshold of matched b/y ions when a given query spectrum matched to the candidate library spectra. the fault number is 3.
5. “Glycopeptide Spectra” is the information of glycopeptide spectra. Currently supported tandem MS Spectral format is “MGF”.
6. “Result” is the file name of the search results.

4. Output

pMatchGlyco outputs 3 files: *.pmlib , *_libcomp.txt (“*” refers to the library file name), and #.txt (“#” refers to the result file name).

(a) “*.pmlib” - library spectral file

Attributes	Meaning
ID	Library spectral ID number
SpectrumName	Spectral name
Sequence	Peptide sequence
Modifications	Modifications occurred at the peptide sequence
PrecursorIonMass	Precursor ion MH+
Charge	Charge of the precursor ion
Frequency	Number of spectra that have the same “Sequence”, “Modifications” and “Charge” in de-glycopeptide spectral identifications.
Peaks	Number of fragment ions
MS/MS	Fragment ions with the format: m/z-intensity-fragment ion number-ion type-amino acid number-charge

Note: “fragment ion number” refers to how many spectra having this fragment ions; “ion type” means ion types (0=b ion, 1=y ion, 2=Y ion, 3=oxonium ion,4=unknown ion); “amino acid number”: how many amino acids of the ion; “charge”: charge of the fragment ion. For example, (532.2514 0.5645 2 0 5 1) means that the m/z of the fragment ion is 532.2514 Da, relative intensity is 0.5645, 2 spectra have this fragment ion, the annotation of this ion is b5+.

(b) “*_libcomp.txt” – library spectral composition

The library spectral composition from de-glycopeptide spectra.

(c) “#.txt” – the glycopeptide identification file

Attributes	Meaning
QuerySpecName	Query spectral name
LibSpecName	Library spectral name
QueryPrecursorMH	Query spectral precursor MH+
LibPrecursorMH	Library spectral precursor MH+
Charge	Query spectral precursor charge
Sequence	Identified peptide sequence
Modification	Identified Modifications
MassShift	QueryPrecursorMH-LibPrecursorMH
GlycanMass	Identified glycan mass
GlycanName	Identified glycan name
GlycoLoc	Glycan location
LibCandidateNum	Number of candidate library spectra
MatchedYIonNum	Number of matched Y ions
MatchedYIonIntSum	Intensity sum of matched Y ions
MatchedPepIonNum	Number of matched b/y ions
MatchedPepIonIntSum	Intensity sum of matched b/y ions
DPS	SDP_Score
PS	P_Score
pMatchGlyco_Score	pMatchGlyco_Score
Target/DecoyHit	Target(T) or Decoy(D) hit
DPSIon	Ions contribute to SDP_Score
PSIon	Ions contribute to P_Score
FDR	False discovery rate

5. FAQ

(1) Q: In the section of “AA” of file “MassTable.ini”, there is a symbol of “J”, what is it?

pMatchGlyco changes amino acid “N” within a peptide sequon of “N-X-S/T/C” to “J”, where X can be any amino acid except “P”, Therefore, the mass of amino acid of “J” also should be set be equal to the mass of amino acid of “N”, i.e. 114.04293 Da.

(2) Q: What the symbol “J” represents in the peptide sequences of library spectra and result files?

pMatchGlyco changes amino acid “N” within a peptide sequon of “N-X-S/T/C” to “J”, where X can be any amino acid except “P”. In “MassTable.ini”, the mass of amino acid of “J” also should be set be equal to the mass of amino acid of “N”.