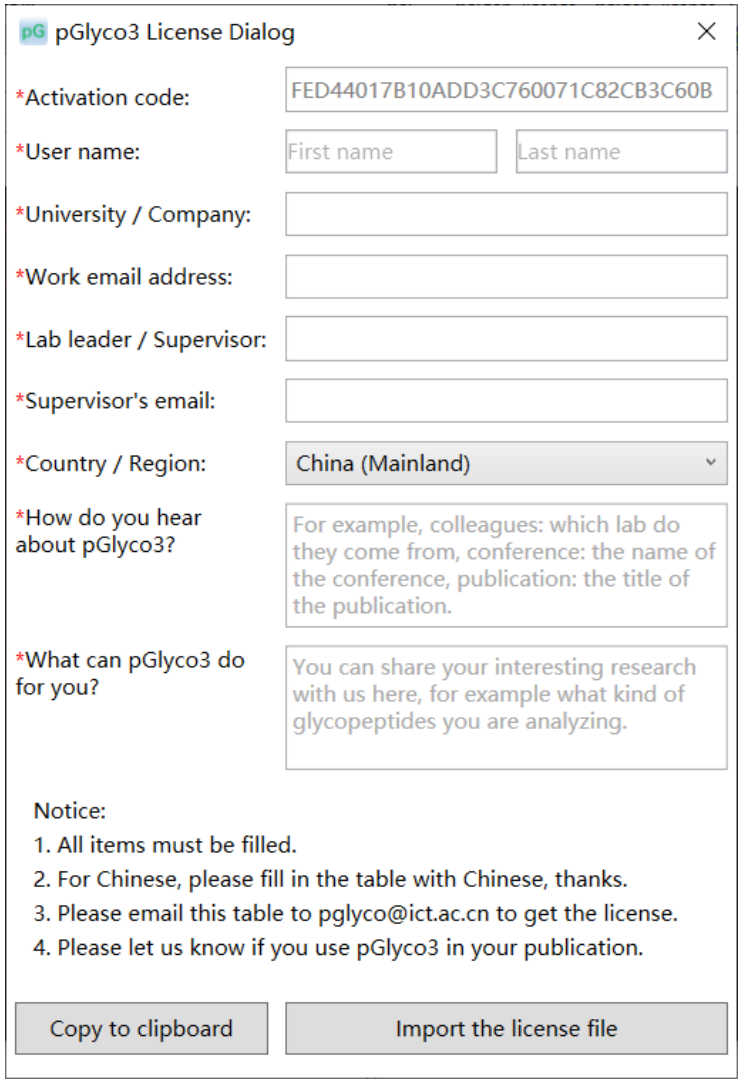
pGlyco3 User Guide

Dr. Wen-Feng Zeng

2021.04.01

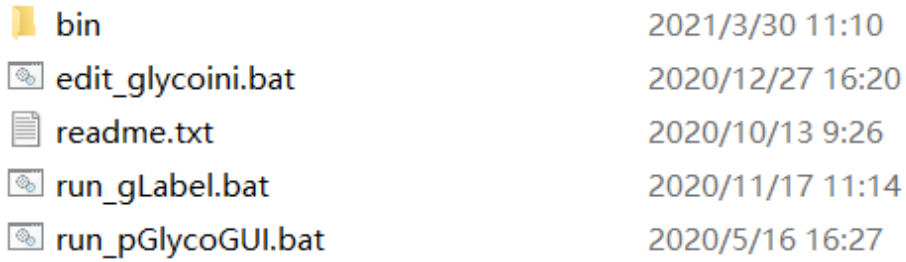
pGlyco3

- Download the latest version from: https://github.com/pFindStudio/pGlyco3/releases
- Complete the form, click “Copy to clipboard”
- Paste the form in http://i.pfind.org/license/pGlyco3
- After receive the pGlyco3.license (check the spam), click “Import the license file” to import it into pGlyco3
- Restart pGlyco3



Runners

- run_pGlycoGUI.bat
  - pGlyco search engine
- run_gLabel.bat
  - Spectrum annotation
- edit_glycoini.bat
  - Add/edit monosaccharides

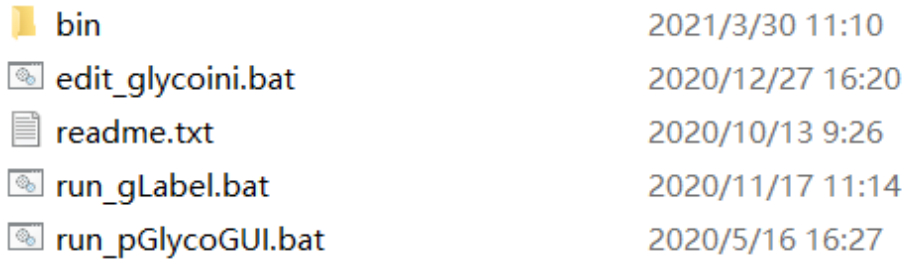


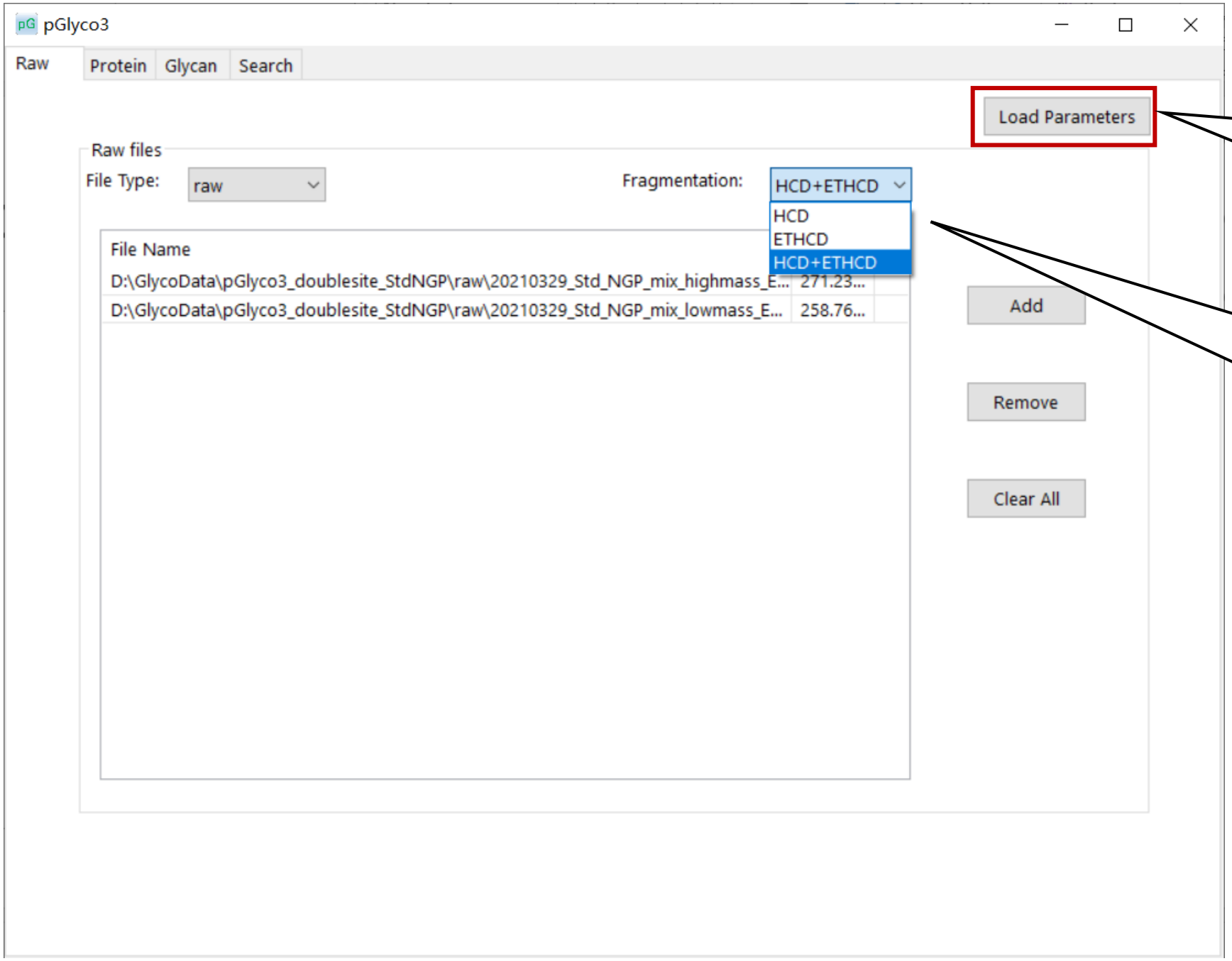
Runners

- run_pGlycoGUI.bat
  - pGlyco search engine

- run_gLabel.bat
  - Spectrum annotation

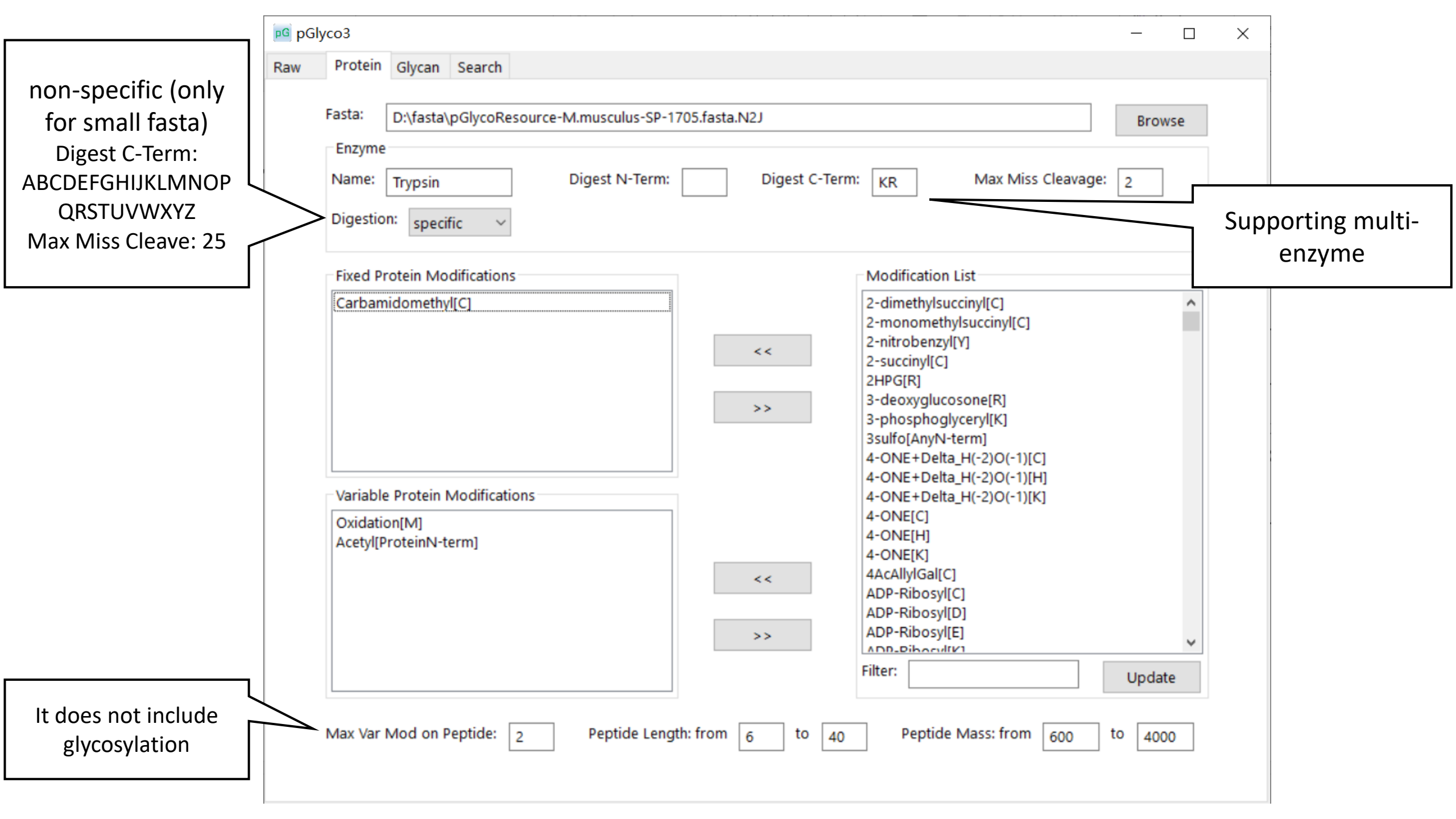
- edit_glycoini.bat
  - Add/edit monosaccharides



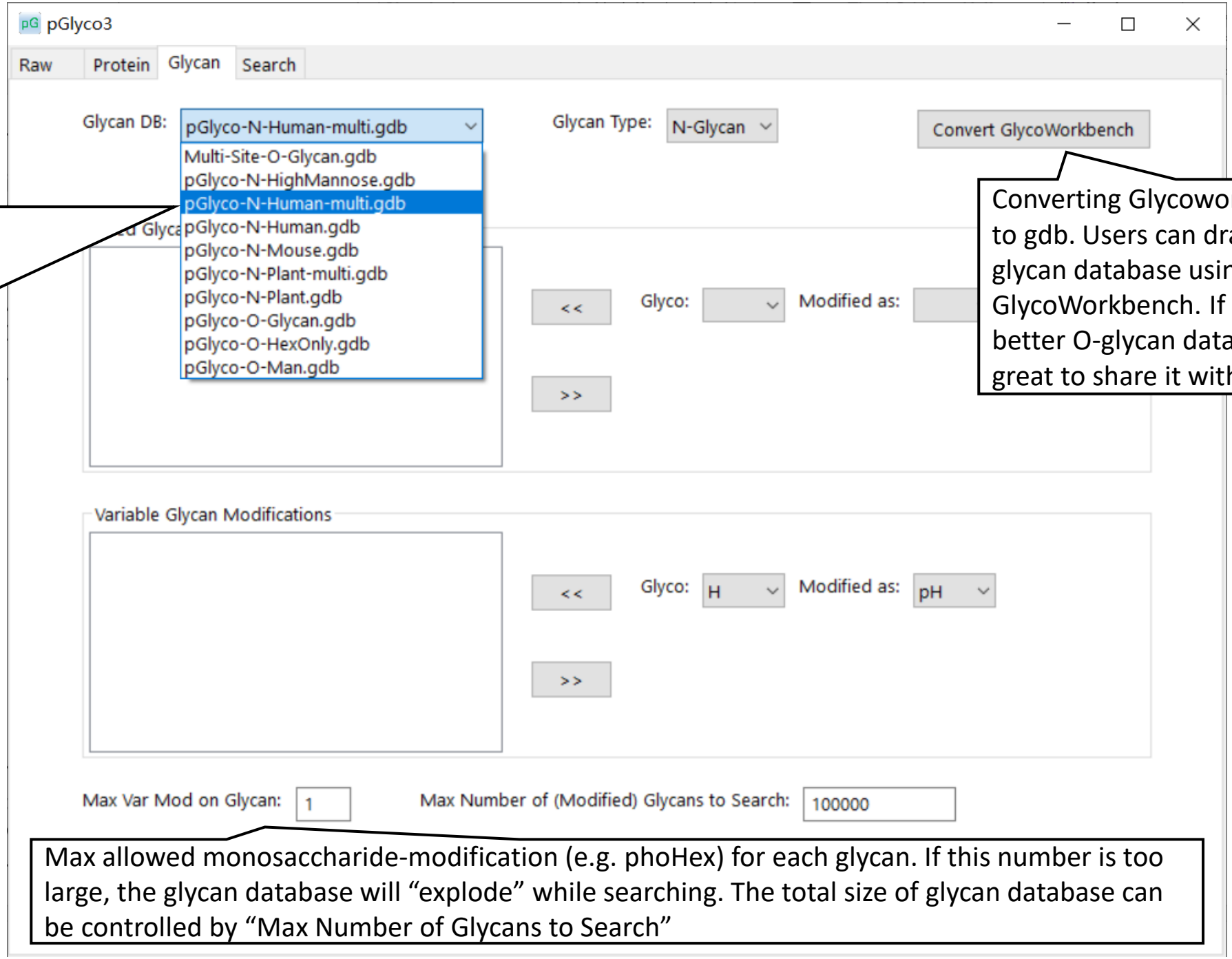


Loading existing parameters (pGlyco.cfg)

HCD+ETHCD could be: HCD-pd-ETxxD, HCD-couple-ETxxD, or even HCD-pd-CID-ETxxD, etc.

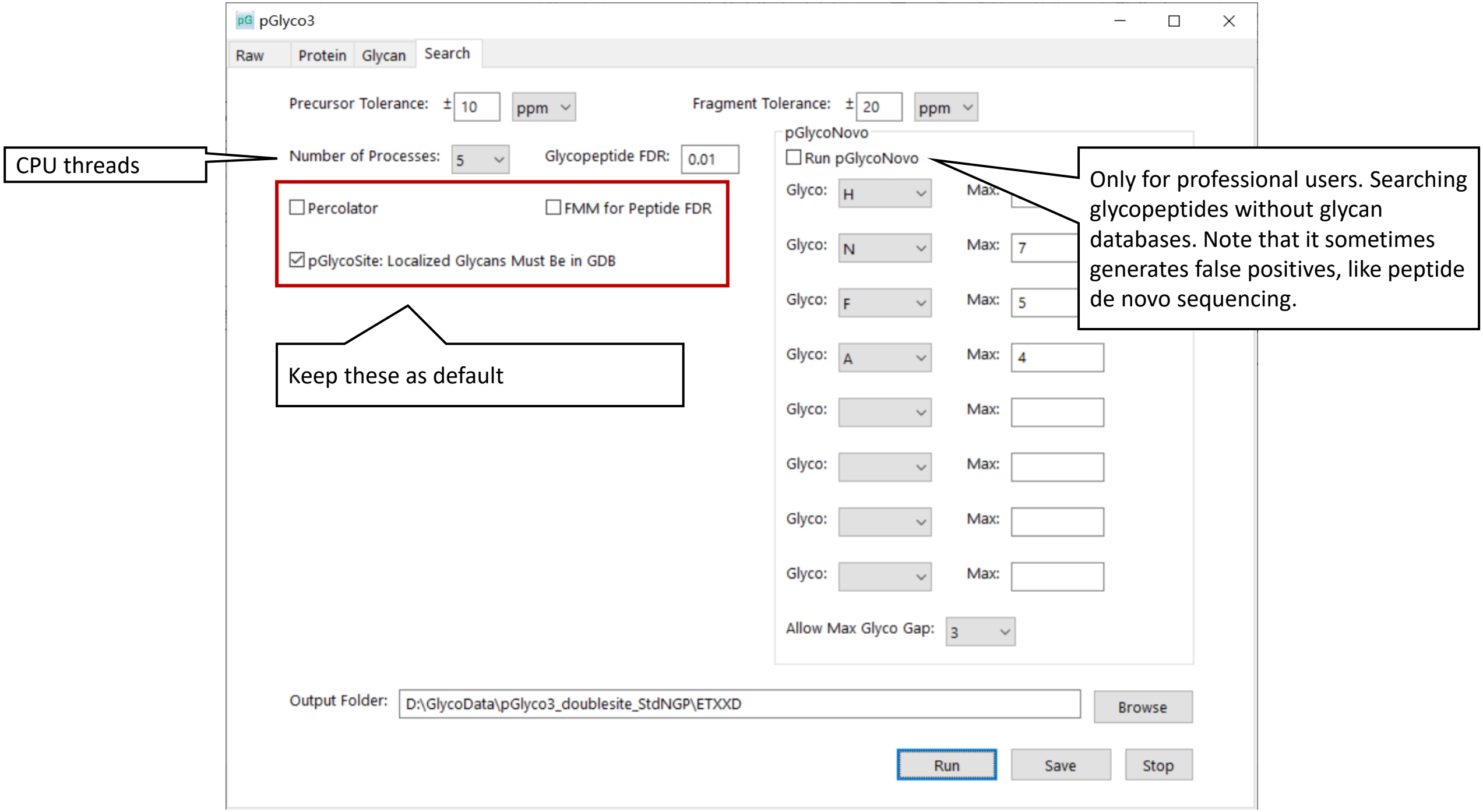

non-specific (only for small fasta)
Digest C-Term: ABCDEFGHIJKLMNOP QRSTUVWXYZ
Max Miss Cleave: 25
pGlyco3
Raw
Protein
Glycan
Search
Fasta:
D:\fasta\pGlycoResource-M.musculus-SP-1705.fasta.N2J
Browse
Enzyme
Name:
Trypsin
Digest N-Term:
Digest C-Term:
KR
Max Miss Cleavage:
2
Digestion:
specific
Supporting multi-enzyme
Fixed Protein Modifications
Carbamidomethyl[C]
<<
>>
Modification List
2-dimethylsuccinyl[C]
2-monomethylsuccinyl[C]
2-nitrobenzyl[Y]
2-succinyl[C]
2HPG[R]
3-deoxyglucosone[R]
3-phosphoglyceryl[K]
3sulfo[AnyN-term]
4-ONE+Delta_H(-2)O(-1)[C]
4-ONE+Delta_H(-2)O(-1)[H]
4-ONE+Delta_H(-2)O(-1)[K]
4-ONE[C]
4-ONE[H]
4-ONE[K]
4AcAllylGal[C]
ADP-Ribosyl[C]
ADP-Ribosyl[D]
ADP-Ribosyl[E]
Variable Protein Modifications
Oxidation[M]
Acetyl[ProteinN-term]
<<
>>
Filter:
Update
It does not include glycosylation
Max Var Mod on Peptide:
2
Peptide Length: from
6
to
40
Peptide Mass: from
600
to
4000

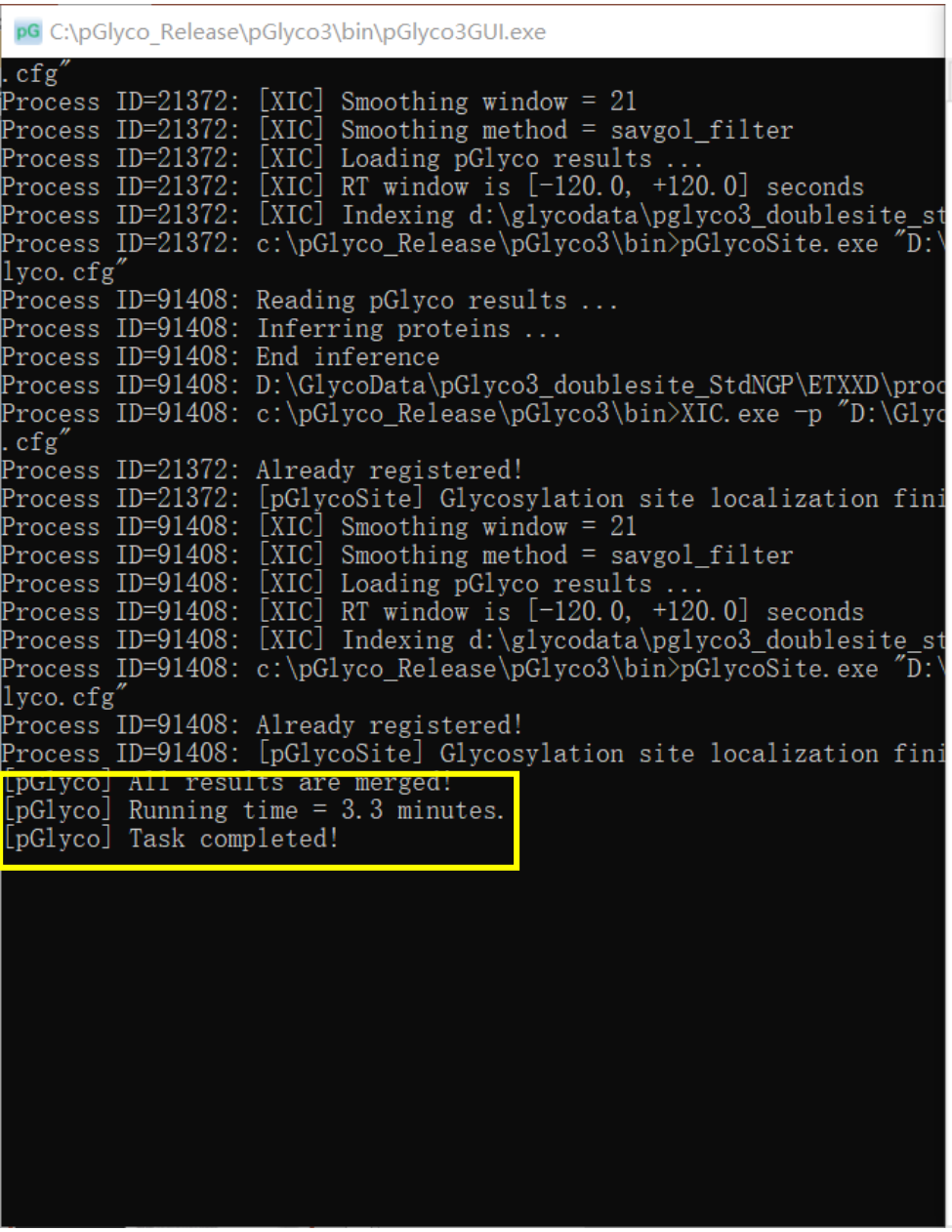
For O-Glycan, Multi-Site-O-Glycan.gdb is recommended. Multi-site N-glycans could also be localized by ETxxD using xxx-multi.gdb

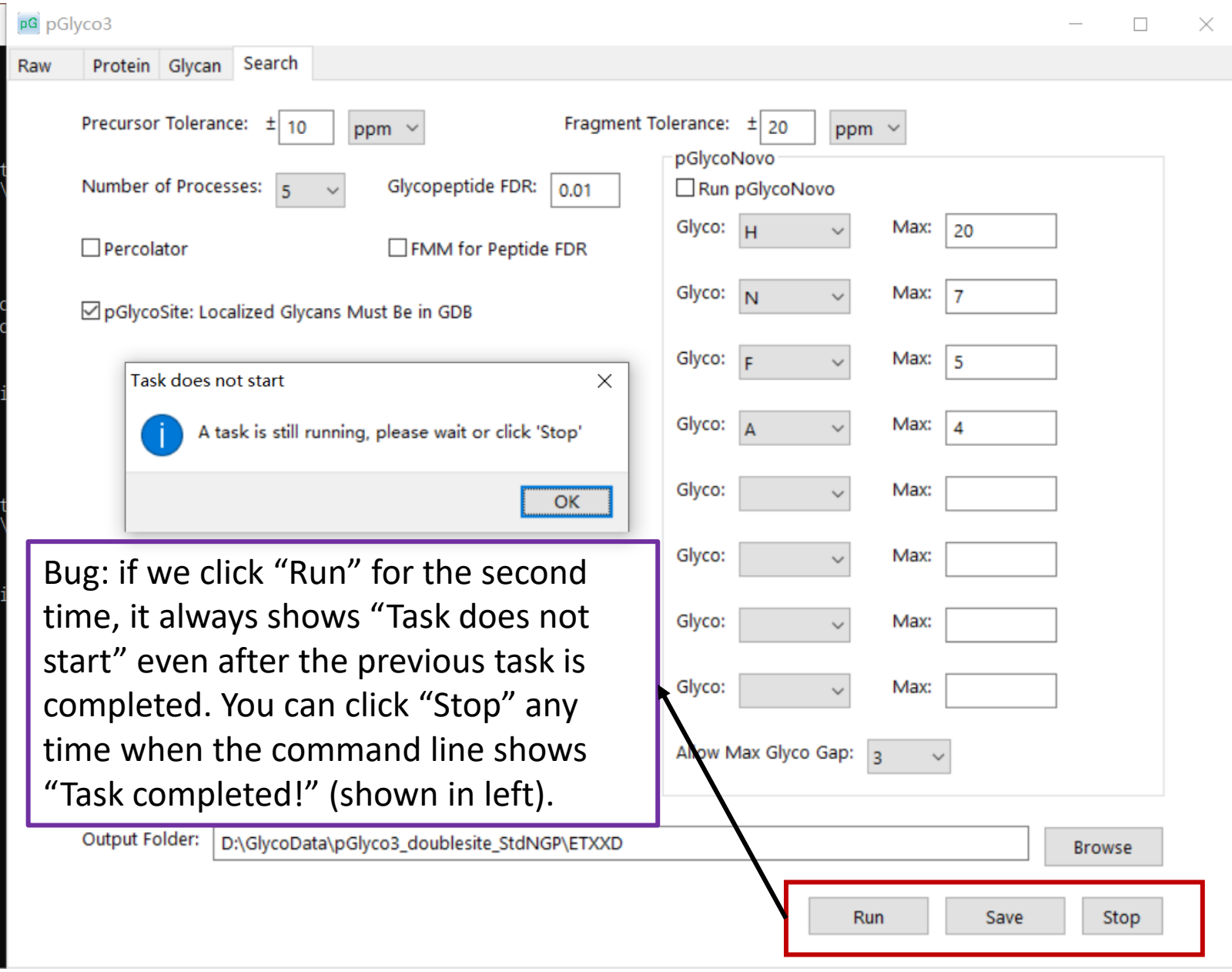


Converting Glycoworkbench .gwp file to gdb. Users can draw their own glycan database using GlycoWorkbench. If you have a better O-glycan database, it will be great to share it with others☺

Max allowed monosaccharide-modification (e.g. phoHex) for each glycan. If this number is too large, the glycan database will “explode” while searching. The total size of glycan database can be controlled by “Max Number of Glycans to Search”


pGlyco3
Raw
Protein
Glycan
Search
Precursor Tolerance: ± 10 ppm
Fragment Tolerance: ± 20 ppm
Number of Processes: 5
Glycopeptide FDR: 0.01
Percolator
FMM for Peptide FDR
pGlycoSite: Localized Glycans Must Be in GDB
pGlycoNovo
Run pGlycoNovo
Glyco: H
Max:
Glyco: N
Max: 7
Glyco: F
Max: 5
Glyco: A
Max: 4
Glyco:
Max:
Glyco:
Max:
Glyco:
Max:
Glyco:
Max:
Allow Max Glyco Gap: 3
Output Folder: D:\GlycoData\pGlyco3_doublesite_StdNGP\ETXXD
Browse
Run
Save
Stop
CPU threads
Keep these as default
Only for professional users. Searching glycopeptides without glycan databases. Note that it sometimes generates false positives, like peptide de novo sequencing.





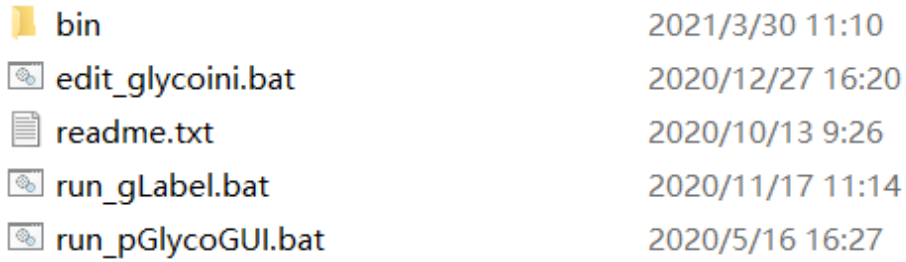
Bug: if we click “Run” for the second time, it always shows “Task does not start” even after the previous task is completed. You can click “Stop” any time when the command line shows “Task completed!” (shown in left).

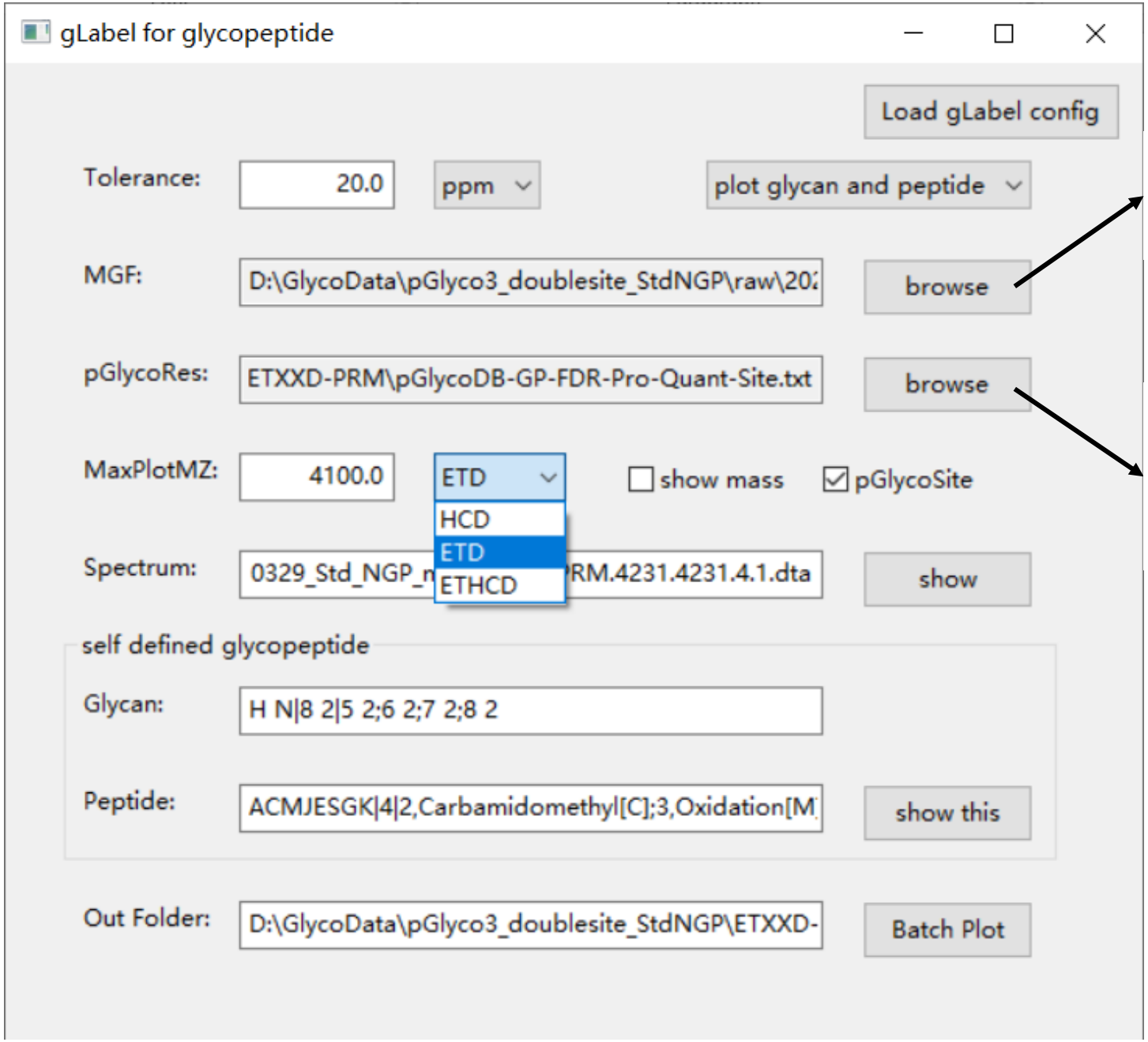
Runners

- run_pGlycoGUI.bat
  - pGlyco search engine

- run_gLabel.bat
  - Spectrum annotation

- edit_glycoini.bat
  - Add/edit monosaccharides





20210329_Std_NGP_mix_ETXXD_PRM_ETHCDFT.mgf **EThcD**

20210329_Std_NGP_mix_ETXXD_PRM_HCDFT.mgf **HCD**

20210329_Std_NGP_mix_ETXXD_PRM_HCDpdETXXDFT.mgf

Do not use it

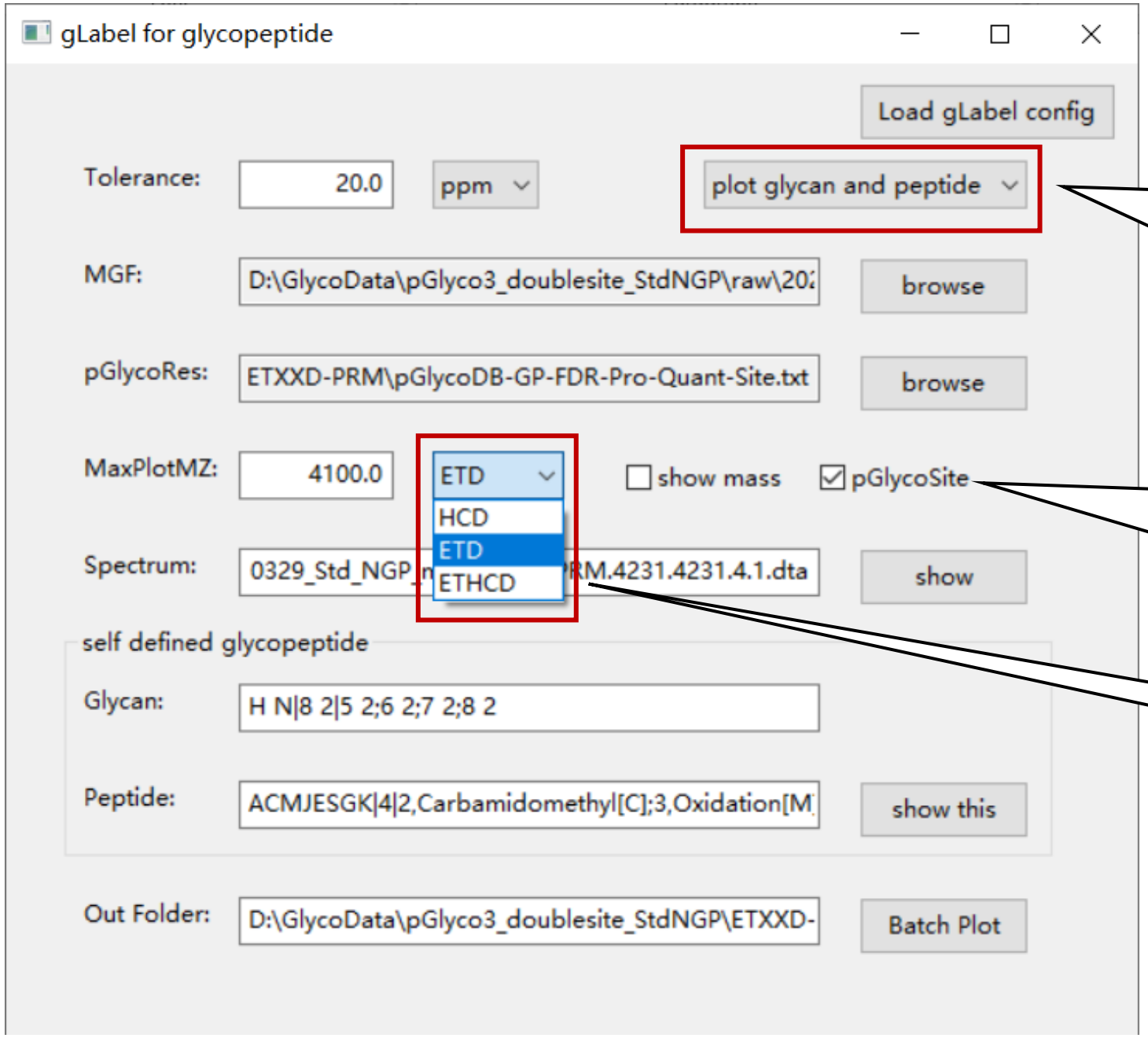
pGlycoDB-GP-FDR.txt

pGlycoDB-GP-FDR-Pro.txt

pGlycoDB-GP-FDR-Pro-Quant.txt

pGlycoDB-GP-FDR-Pro-Quant-Site.txt

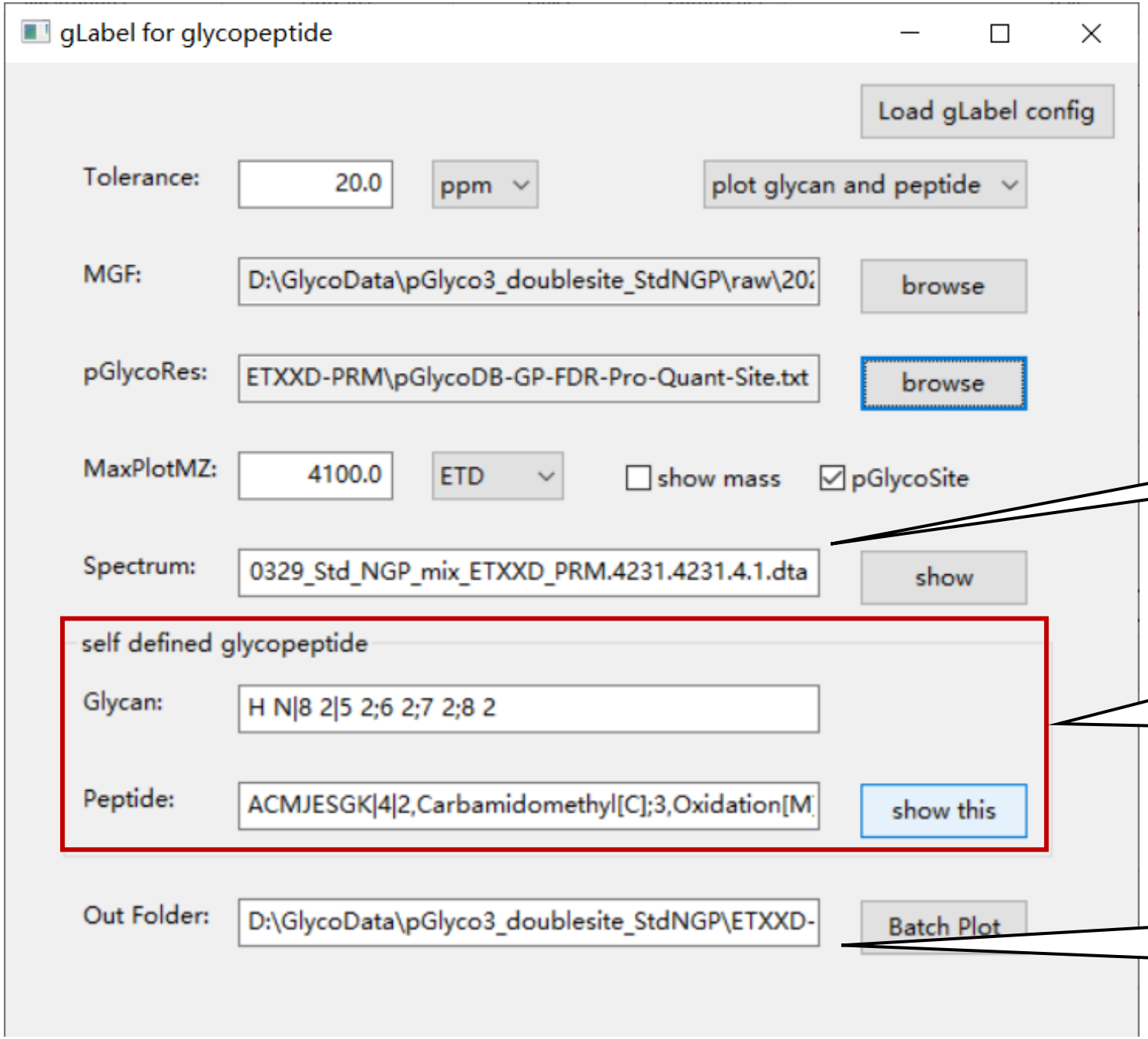
If not sure, choose the longest one



plot glycan: Y ions (with peptide attached)
plot peptide: b/y/c/z ions
plot glycan and peptide: both
B ions are always plotted

Plot glycan-localized c/z ions, only works for ETD mgf

HCD: b/y ions
ETD: c/z ions
ETHCD: b/y/c/z ions



A
GlySpec
20210329_Std_NGP_mix_ETXXD_PRM.3738.3738.4.8.dta
20210329_Std_NGP_mix_ETXXD_PRM.3792.3792.4.7.dta
20210329_Std_NGP_mix_ETXXD_PRM.3837.3837.4.1.dta
20210329_Std_NGP_mix_ETXXD_PRM.3972.3972.4.3.dta
20210329_Std_NGP_mix_ETXXD_PRM.3990.3990.4.1.dta
20210329_Std_NGP_mix_ETXXD_PRM.4035.4035.4.3.dta
20210329_Std_NGP_mix_ETXXD_PRM.4043.4043.3.2.dta
20210329_Std_NGP_mix_ETXXD_PRM.4044.4044.4.1.dta
20210329_Std_NGP_mix_ETXXD_PRM.4052.4052.3.4.dta

First column in the pGlycoRes file, it will automatically switch between HCD scan and ETD scan.

Ignore this section!!

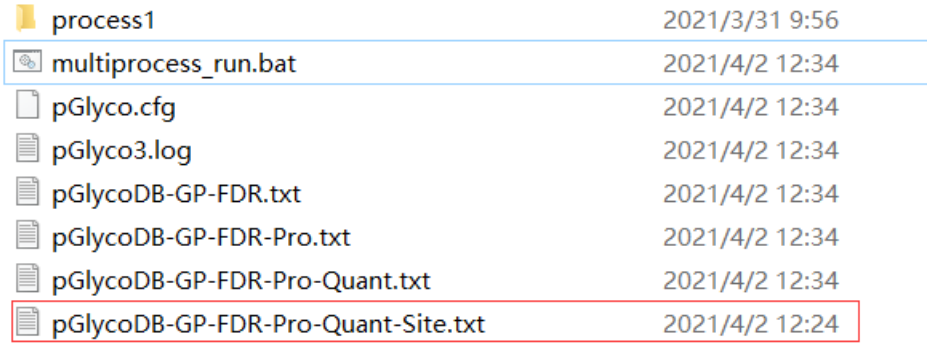
Save all annotated plots into a folder, for lazy people like me.

Result Interpretation

- MonoArea:
  - Monoisotopic peak intensity
- IsotopeArea:
  - Summed intensity of all isotopes (including Mono)
- ETDScan: ETDScan for site localization

- LocalizedSiteGroups:
  - Format: site1,site2,glycan composition,probability; s1,s2,glycan,prob;…

  - if site1≠site2, it means ETD spectrum is not able to distinguish all sites between site1 and site2, site1-to-site2 is regarded as a site-group



AN	AO	AP	AQ	AR	AS	AT
MonoArea	IsotopeArea	ETDScan	LocalizedSiteGroups	LocalizedS	LocalizedI	PreLocalized
87948.9	596593	3742	J9,J9,(5 4 0 0),0.57	3	0.083333	3
591603	4.34E+06	3796	J3,J9,(5 4 0 0),0.91	14	0.388889	14
5.45E+06	5.04E+07	3841	J9,J9,(5 4 0 0),0.85	17	0.472222	15
965899	6.38E+06	3976	J3,J3,(5 4 0 0),0.94	20	0.555556	20
8.22E+06	1.32E+08	3994	J3,J3,(5 4 0 0),0.97	24	0.666667	24
562654	5.09E+07	4039	J3,J9,(5 4 0 0),0.81	7	0.194444	7
700256	1.50E+07	4047	J3,J9,(5 4 0 0),0.81	3	0.083333	3
562654	5.09E+07	4048	J9,J9,(5 4 0 0),0.88	15	0.416667	15
700256	1.50E+07	4056	J3,J3,(5 4 0 0),0.80	6	0.166667	6
562654	5.09E+07	4057	J9,J9,(5 4 0 0),0.90	21	0.583333	21
700256	1.50E+07	4065	J3,J9,(5 4 0 0),0.60	2	0.055556	2
562654	5.09E+07	4066	J3,J3,(5 4 0 0),0.89	11	0.305556	11
562654	5.09E+07	4075	J3,J9,(5 4 0 0),0.73	5	0.138889	5

Python Codes for Post Analysis

- https://github.com/pFindStudio/pGlyco3/blob/main/protein_site_analysis.py
  - Generate protein level information (localized sites, intensities)

- Usage: python protein_site_analysis.py C:/xx/output_dir/pGlycoDB-GP-FDR-Pro-Quant-Site.txt

- It will generate a “pGlyco3-ProSite-Analysis.txt” file in the same folder as pGlycoDB-GP-FDR-Pro-Quant-Site.txt

pGlyco3-ProSite-Analysis.txt

Sites on Protein Sequence

Suggested Cut-off:
0.75 (or 0.8)

Unique Site or Site-group

A	B	C	D	E	F	G	H	I	J
Protein	Gene	ProteinSiteGroup	StartSite	EndSite	LocalizedGlycan(H,N,A,F)	SiteProbability	IsUniqueSite	IsGroupSite	PeptideSiteGroup
sp|P00748	F12	T297:T297	297	297	1 1 1 0	0.89	1	0	T13:T13
sp|P00748	F12	T305:T305	305	305	2 2 2 0	0.85	1	0	T21:T21
sp|P00748	F12	S308:S308	308	308	1 1 1 0	0.9	1	0	S24:S24
sp|P00748	F12	S308:S308	308	308	1 1 1 0	0.89	1	0	S24:S24
sp|P00748	F12	S308:S308	308	308	1 1 1 0	0.88	1	0	S24:S24
sp|P00748	F12	S308:S308	308	308	1 1 1 0	0.85	1	0	S24:S24
sp|P00748	F12	S308:S308	308	308	2 2 1 0	0.8	1	0	S24:S24
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.97	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.97	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.97	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.94	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.94	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.94	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.93	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.93	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.92	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.92	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.92	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.92	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.91	1	0	T18:T18
sp|P00748	F12	T328:T328	328	328	1 1 1 0	0.91	1	0	T18:T18

Python Codes for Post Analysis

- https://github.com/pFindStudio/pGlyco3/blob/main/Y_ion_extractor.py
  - Extract Y ions, e.g. Y-N(1)F(1) and Y-N(2)F(1) to check core fucosylation
  - Y-N(1)F(1) and Y-N(2)F(1) could be modified in the codes

- Usage: python Y_ion_extractor.py C:/xx/ourput_dir/pGlyco3.cfg

We will upload more Python codes

- See https://github.com/pFindStudio/pGlyco3

Advanced usages: to be continued