User Manual for pChem

Version 1.0

Contents

1. Requirements

1) Computing system

pChem search requires a computer with recommended configuration as follows:

- Microsoft Windows 64-bit
- Intel Core i7/i9/Xeon Processor
- 32GB of RAM or more

Note: pChem v1.0 is NOT supported by non-Windows operating systems (incl. MacOS, Linux and so on).

2) MS Data

- Data dependent acquisition (DDA) with BOTH MS1 and MS/MS spectra recorded in the High-Resolution mode

Note: 1) For automatic performance assessment of chemoproteomic probes, it is recommended to acquire MS data from probe-labeled samples with isotope-coding. 2) MS data from non-isotope-labeled samples can also be processed by pChem, but the search result may be subjected to manual inspection.

2. Download

1) pChem can be freely downloaded from the website: http://pfind.org/software/pChem/index.html

pFind Studio: a computational solution for mass spectrometry-based proteomics

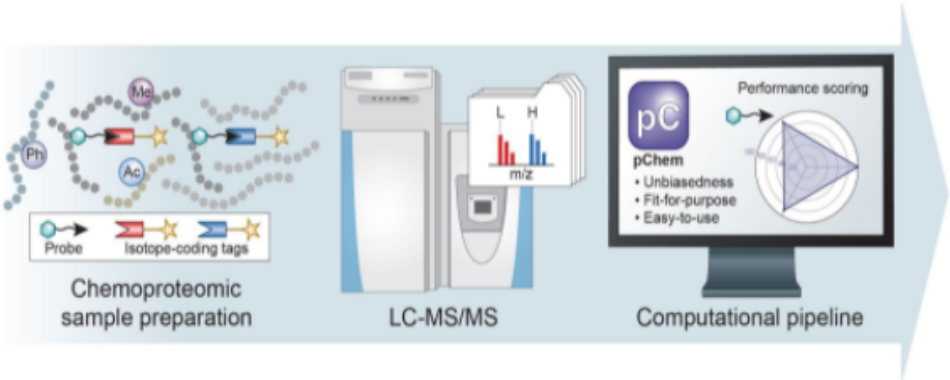
Home Members Publications Applications Software Links About us

pChem

Introduction - Cite us - Downloads

Introduction

Chemical probe coupled with mass spectrometry (MS)-based proteomics, herein termed chemoproteomics, offers versatile tools to globally profile protein features and to systematically interrogate the mode of action of small molecules in a native biological system. Nonetheless, development of an efficient and selective probe for chemoproteomics can still be challenging. Besides, it is also difficult to unbiasedly assess its chemoselectivity at a proteome-wide scale. Here we present pChem, a modification-centric blind search and summarization tool to provide a pipeline for rapid and unbiased assessing of the performance of ABPP and metabolic labeling probes. This pipeline starts experimentally by isotopic coding of PDMs, which can be automatically recognized, paired, and accurately reported by pChem, further allowing users to score the profiling efficiency, modification-homogeneity and proteome-wide residue selectivity of a chemoproteomic probe.



Cite us

pChem: a modification-centric assessment tool for performance of chemoproteomic probes.
Ji-Xiang He, Zheng-Cong Fei, Fu-Chu He, Si-Min He, Hao Chi, Jing Yang.
Under review

Downloads

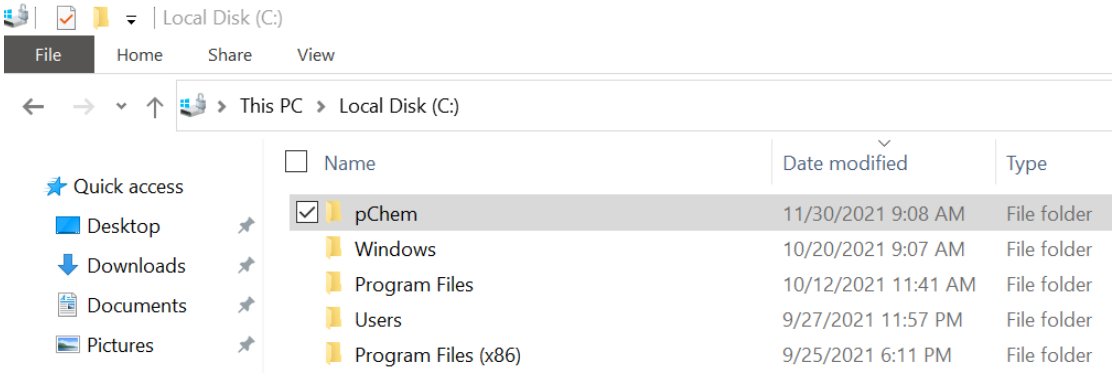
pChem version 1.0 is currently free to use. click to download.

For source code, please refer to github.

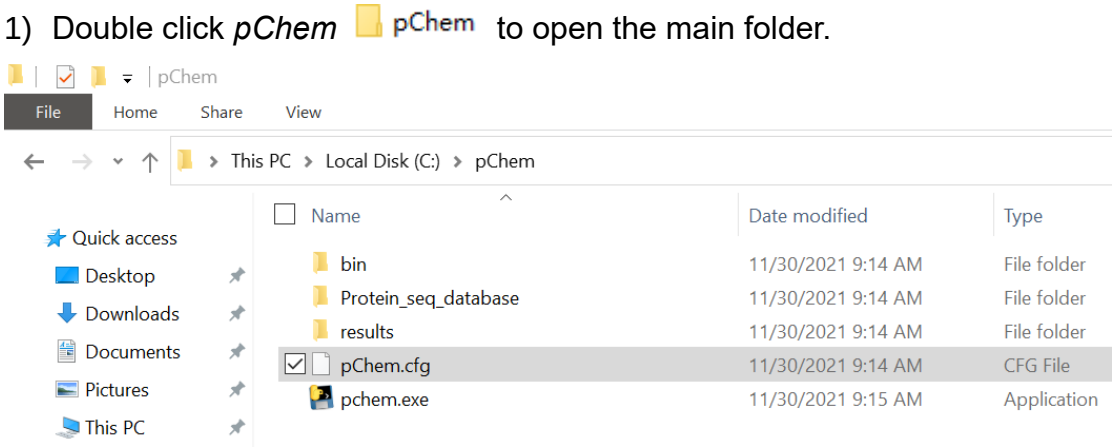
For detailed usage, please refer to user guide.

2) click “*click to download*” to get the zipped software.

3) Un-zip the “pChem.zip” package into a specified file folder (e.g., Local disk C).



3. Configuration

1) Double click *pChem* 

to open the main folder.

2) Open configuration file "*pChem.cfg*" using a text editor, *e.g.*, Microsoft Notepad or Notepad++ (https://notepad-plus.en.softonic.com/).

3) Setting "pChem.cfg".

```
# If isotope coding is adopted to facilitate the discovery of unknown modifications (True or False)
isotope_labeling=True

# Path to the output file
output_path=D:\pchem\pChem_new\results

# Path to the protein sequence database
fasta_path=D:\pChem\pChem_new\Protein_seq_database\Homo_sapiens_uniprot_canonical_20395_entries_20210516.fasta

# Format of MS data, RAW or MZML
msmstype=RAW

# The number and path of MS data
msmsnum=1
msmspath1=D:\pchem\pChem_data\QE_Plus_YangJing_FL_ALK_50per_20170531.raw


# Type of MS dissociation method
activation_type=HCD-FTMS

# Usage of open search (True/ False), against Unimod, the common modification can be set if not
open_flag=False
common_modification_number=2
common_modification_list=Carbamidomethyl[C];Oxidation[M];


# Mass tolerance of the mass shift between light isotope and heavy isotope
mass_of_diff_diff=6.020132


# Isotopic mass difference within empirically defined tolerance(Da)
mass_diff_diff_range=0.005

# Mass range of unknown modification (Da)
min_mass_modification=200
max_mass_modification=1000

# Isotopic pairs of mass shifts with PSMs less than X% of that of overall PDMs were neglected
filter_frequency=5

# If consider the N-side or C-side for amino acid localization (True or False)
side_position=True

# P-value threshold enabling confident amino acid localization
p_value_threshold=0.001

# If report the statistical information (True or False)
report_statistics=False
```

General Note 1:
For the first-time users, custom settings are required for ①-⑤, ⑧ default settings can be adopted for ⑥, ⑦, ⑨-⑭.
General Note 2:
All parameters (shown in red below) are case sensitive.
General Note 3:
The blank space should be avoided.

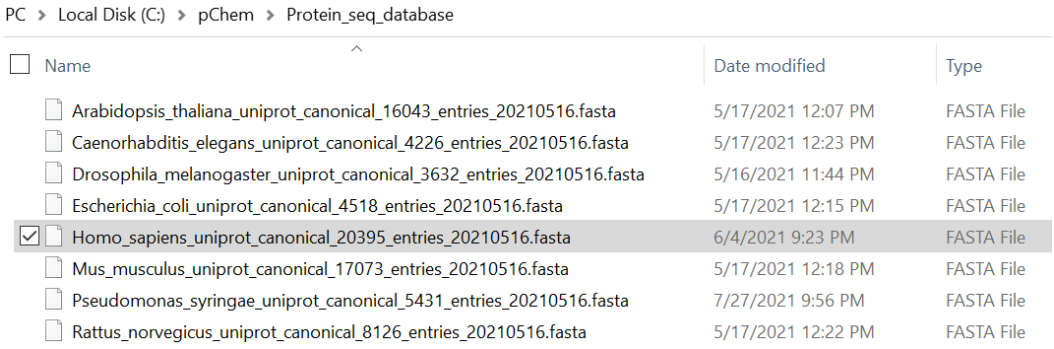
① # If isotope coding is adopted to facilitate the discovery of unknown modifications (True or False)
Isotope_labeling=True
illustration: default
Note: Choose 'False', if pChem is adopted to search endogenous modifications from probe-free and/or label-free protein samples,

② # Path to the output file

output_path=C:\pChem\results
Note: If the output file folder does not exist, an error will be reported.

③ # Path to the protein sequence database

fasta_path=C:\pChem\Protein_seq_database\Homo_sapiens_uniprot_canonical_20395_entries_20210516.fasta
Note: The protein *.fasta database databases of several commonly used species (*e.g.*, *home sapiens*) are included in the subfolder (named as Protein_seq_database) of pChem. Note that the databases of other species can be downloaded from Uniprot as described in **Supporting Protocol 1**.



④ # Format of MS data (RAW or MZML)

msmstype=RAW
Note: Non-Thermo MS data need to be converted into mzML files before pChem search. The users can refer to **Supporting Protocol 2**.

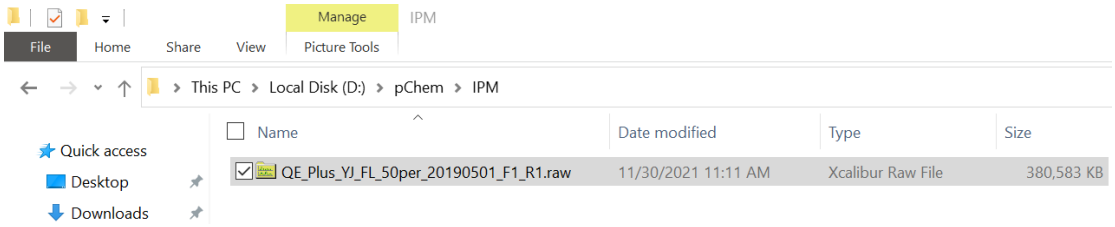
⑤ # The number and path of MS data

msmsnum=N
msmspath1=X:\XXX\XXX.raw
msmspath2=X:\XXX\XXX.raw
……
msmspathN=X:\XXX\XXX.raw

Note: The suffix of MS data files MUST be input.

Example:

msmsnum=1
msmspath1=D:\pChem\IPM\QE_Plus_YJ_FL_50per_20190501_F1_R1.raw



⑥ # Type of MS dissociation method

activation_type=HCD-FTMS

illustration: default

Note: 1) Users can adopt this default setting if their MS data is generated by TOF (Time-of-flight) instruments implementing CID (Collision-induced dissociation)-type of fragmentation (e.g., SCIEX 5600, SCIEX 6600, Bruker TimsTOF); 2) pChem v1.0 can NOT support MS data generated under electron-transfer dissociation ETD, electron-transfer/higher-energy collision dissociation EThcD, and the likes.

⑦ # Usage of open search (True/ False) against Unimod, the common modification can be set if not

open_flag=False
common_modification_number=2
common_modification_list=Carbamidomethyl[C];Oxidation[M];

illustration: default

Note: The names of common modifications should be the same as those appeared in Unimod database.

⑧ # Mass tolerance of the mass shift between light isotope and heavy isotope

mass_of_diff_diff=6.020132

Note: This default value is calculated based on the isotopic mass shift between six heavy and light carbons encoded in probe-derived modifications (PDMs). Users can set any other values based on their different isotope labeling strategies.

Troubleshooting: One needs to confirm this value being correctly input.

⑨ # Isotopic mass difference within empirically defined tolerance (Da)

mass_diff_diff_range=0.005
illustration: default
Troubleshooting: If the pChem search mis-identified the targeted PDMs or even report nothing, one might want to loose the defined mass tolerance (e.g., 0.01Da).

⑩ # Mass range of unknown modification (Da)

min_mass_modification=200
max_mass_modification=1000
illustration: default
Note: The PDMs generated from the use of bioorthogonal cleavable linkers typically possess masses higher than 200 Da and less than 1000Da.

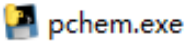
⑪ # Isotopic pairs of mass shifts with PSMs less than X% of that of overall PDMs were neglected
filter_frequency=5
illustration: default
Note: This parameter can be set as 0 if one wants to retrieve all PDMs including those with just a few PSMs.

⑫ # If consider the N- or C-termini for amino acid localization (True or False)
side_position=True
illustration: default

⑬ # P-value threshold enabling confident amino acid localization
p_value_ threshold=0.001
illustration: default

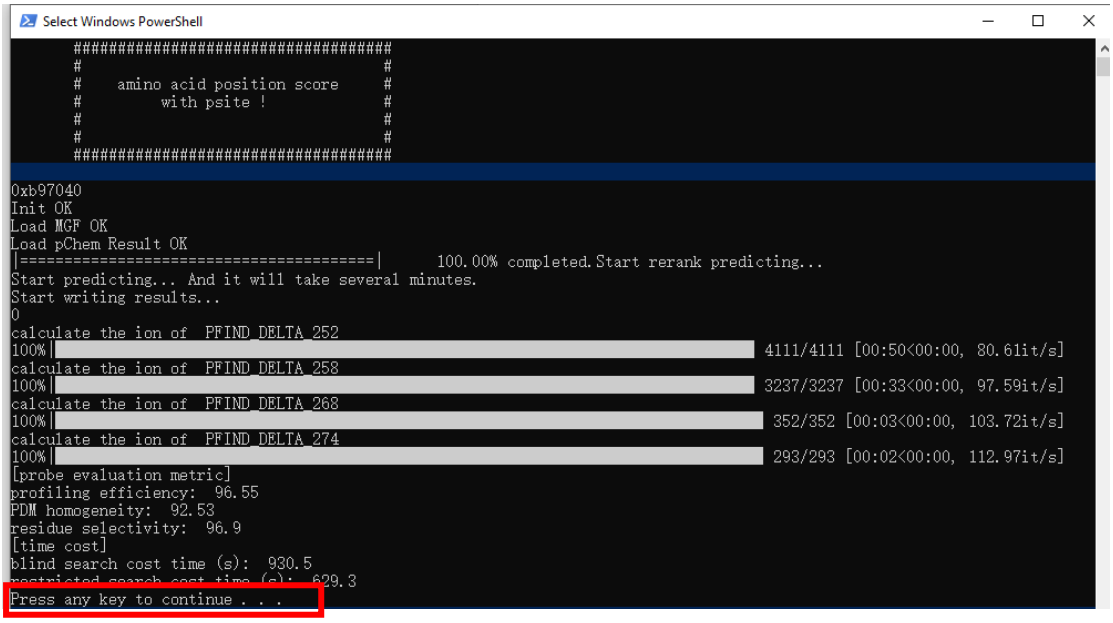
⑭ # if report the statistical information (True or False)
report_statistics=False
illustration: default

4. Run

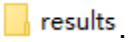
Once all parameters have been set, double click "*pChem.exe*" 

to execute the programming. The message "**Please press any key to continue**" means that program runs to completion.

Note: pChem search will generate several intermediate files in the main folder. do NOT open those files during program running.

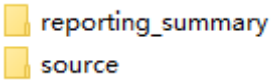


5. Output

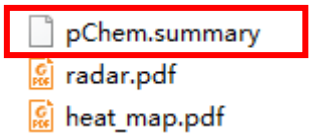
1) Double click "*results*" file for searching results 

.

2) Double click "*reporting summary*".



3) There are three major output documents.



Note: Users are recommended to copy these output documents and paste into another file. Otherwise, they can be covered by those generated from the next search event.

① pChem.summary

pChem.summary is a tab-delimited text file contains the details of every PDM.

Rank	PDM	Accurate Mass	Top1 Site|Probability	Others	#PSM	#PSM L|H	DFLs
1	PFIND_DELTA_252	252.122339	C|0.988		13876	7368|6508	
2	PFIND_DELTA_268	268.116411	C|0.487	M(0.291); N-SIDE(0.212);	1578	872|706	302.104737, 301.103528, 320.113184

PDM: Probe-derived modifications

#PSM: The number of PSMs corresponding to modified peptides identified by targeted search

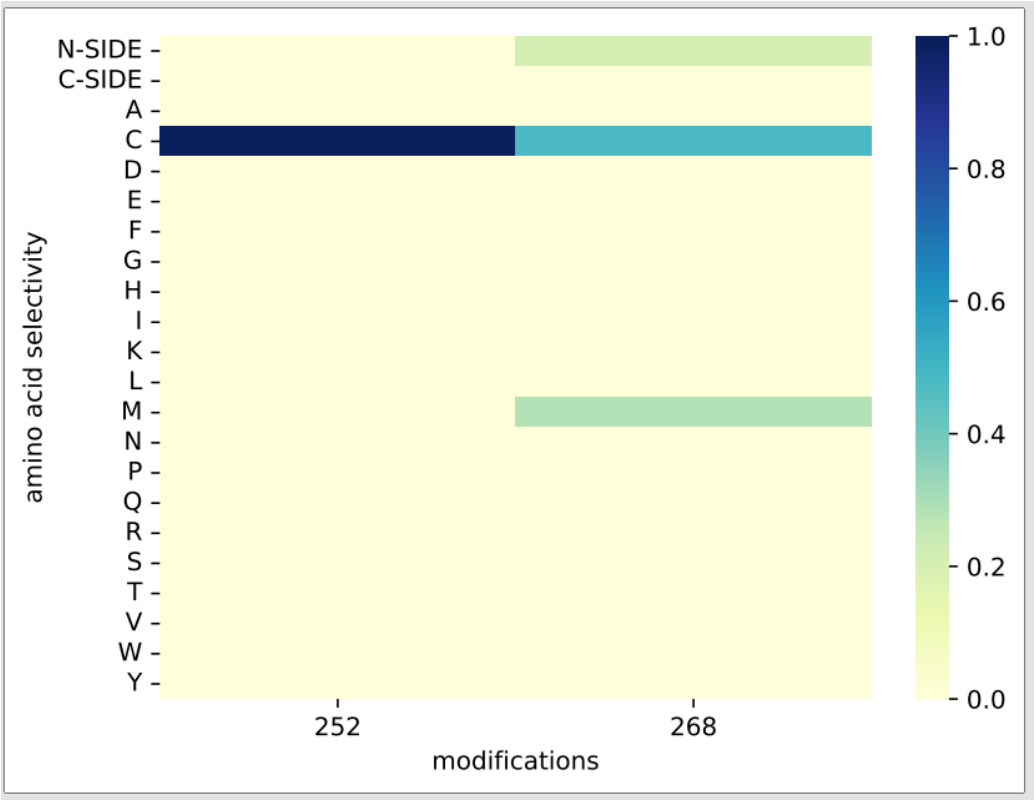
#PSM L|H: The number of PSMs assigning to light and heavy bearing the corresponding PDM, respectively
Note: For data from on-isotope-labeled samples, this information will NOT be shown.

Top1 site | Top1 Probability: The amino acid most likely to be modified with the corresponding localization probability.

Others: Other amino acid sites that may also be labeled by probes and their corresponding localization probability values

DFLs: Diagnostic fragment losses
Note: For data from on-isotope-labeled samples, DFLs will NOT be provided.

② Heat_map.pdf



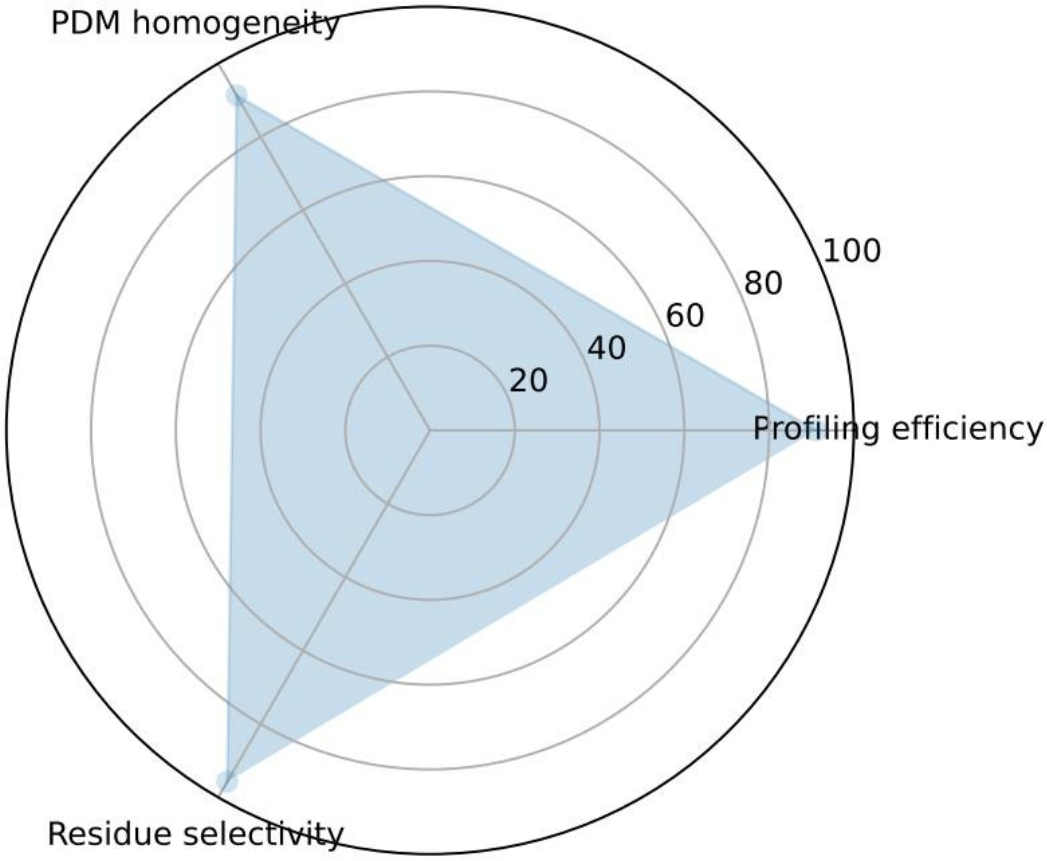
Horizontal coordinate: The Δmass of each PDM

Longitudinal coordinate: The types of amino acids

Color gradient: The localization probability that the modification occurs at each potential site.

Note: 1) Those amino acids with p-value higher than 0.001 are considered mis-localized sites. As such, their localization probability values are defined to be null. 2) For data generated from non-isotope-labeled samples, heatmap will NOT be provided.

③ **Radar.pdf**



Radar.pdf contains a radar plot whose radial axes correspond to the three scores as indicated.

Profiling efficiency *(%)* that evaluates whether a probe enables the efficient identification of modified peptides for chemoproteomics.

Modification homogeneity *(%)* that evaluates whether a probe forms a uniform modification.

Residue selectivity *(%)* that evaluates whether a probe selectively targets specific amino acid:

Note: 1) *PDM homogeneity* and *Residue selectivity* are calculated based on the blind search results, while *Profiling efficiency* is calculated according to restricted search; 2) For data generated from on-isotope-labeled samples, radar plot will NOT be provided.

④ PSM-level results

C:\pChem\results\source\blind\pFind-Filtered.spectra

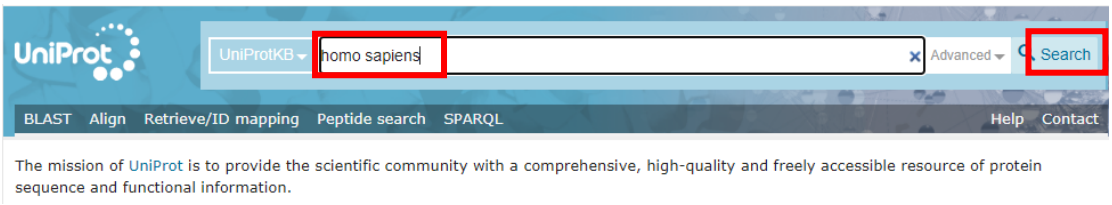
This file contains the PSM-level information regarding all possible modifications on peptides.

1	File_Name	Scan_No	Exp.MH+	Charge	Q-value	Sequence	C	c.MH+	Mass_Shift	Raw_Score	Final_Sc	Modification	Specificity	Proteins	Position	Label	Target/Dec	Miss.Cl	Sin	Avg.Frag	Mass.Sh	Others	Accurate modification mass
2	QE_Plus_Yangling_TCI	34997	3323.517	3		0 NLNDQVLRDQGNRPLPEDMTDSDCR		3323.503	0.013332	32.11487	7.058	2; 20.Oxidation(M);25.PFIND_DELTA_252.12;	3	spiQ14116|IL18_HUI 49 R.D/	1	0|0|PFIND_I	target	1	4.88814	96	252.120478
3	QE_Plus_Yl_FL_50per_	19072	3450.497	3		0 SSSQPSSCCSDPSKPGGNVEGATQSLAEQ		3450.502	-0.00423	33.12794	1.515	0; 8.PFIND_DELTA_309.15;30.Oxidation(M);	3	spiQ13501|SQSTM_I 281 K.K/	1	PFIND_D	target	1	-3.22096	36	309.192429
4	QE_Plus_Yangling_TCI	34275	3198.554	3		0 ASSDLSIASSEEDHLSQNACILESVSEK		3198.549	0.005356	26.01278	2.798	0; 20.PFIND_DELTA_258.15;	3	spiO9H2G2|SLK_HU 338 R.T/	1	PFIND_D	target	1	-0.65771	36	258.142982
5	QE_Plus_Yangling_TCI	34834	3458.562	3		0 CPEAPQPSPLGAIESCGHHETTPNSMK		3458.55	0.011821	33.4101	2.955	0; 1.PFIND_DELTA_252.12;13.Oxidation(M);1	3	spiP60709|ACTB_HU 256 R.C/	1	PFIND_D	target	0	2.01362	100	252.118449
6	QE_Plus_Yangling_TCI	34272	3192.535	3		0 ASSDLSIASSEEDHLSQNACILESVSEK		3192.529	0.005925	25.48116	2.966	0; 20.PFIND_DELTA_252.13;	3	spiQ9H2G2|SLK_HU 338 R.T/	1	PFIND_D	target	1	-1.67882	36	252.123574
7	QE_Plus_Yangling_TCI	25059	2666.205	3		0 VVMALGDYMGASCHACIGGTNVR		2666.192	0.01266	30.79532	6.015	0; 3.Oxidation(M);9.Oxidation(M);13.Carbam	3	spiP60842|IF4A1_HU 118 K.A/	1	0|0|0|PFRI	target	0	2.26348	96	252.122352
8	QE_Plus_Yangling_TCI	28056	2345.168	2		0 KPTDGASSSNCVTDISHLVR		2345.168	-0.00077	30.04682	6.306	0; 11.PFIND_DELTA_258.15;	3	spiP49321|NASP_HU 697 R.K/	1	PFIND_D	target	1	-0.94686	32	258.140161
9	QE_Plus_Yangling_TCI	30373	2684.348	3		0 ILGLQVQQHEHC9QDAQAAMR		2684.351	-0.00316	31.96724	7.855	0; 12.PFIND_DELTA_258.16;21.Oxidation(M)	3	spiQ99ZR2|REXO4_I 370 K.L/	1	PFIND_D	target	0	-2.18117	32	258.146453
10	QE_Plus_Yangling_TCI	35575	2495.261	3		0 LLDVLSGHEGPISGLCRNPMK		2495.25	0.001428	28.07684	9.108	0; 16.PFIND_DELTA_252.13;20.Oxidation(M);	3	spiQ15269|PWP2_HI 492 R.S/	1	PFIND_D	target	0	-2.2878	32	252.121778
11	QE_Plus_Yangling_TCI	25045	2672.221	3		0 VVMALGDYMGASCHACIGGTNVR		2672.212	0.009384	29.89181	1.025	0; 3.Oxidation(M);9.Oxidation(M);13.Carbam	3	spiP60842|IF4A1_HU 118 K.A/	1	0|0|0|PFRI	target	0	3.11740	96	258.139053
12	QE_Plus_Yangling_TCI	22491	2908.322	3		0 HPQVASTWNEVHGHSNAQCVSHNR		2908.312	0.010883	25.1849	1.036	0; 19.PFIND_DELTA_252.12;	3	spiQ15154|PCM1_HI 549 R.D/	1	PFIND_D	target	0	2.5937	100	252.119635
13	QE_Plus_Yangling_TCI	25692	2993.364	3		0 ELPQTPGHTEESMTDDNITEVSCK		2993.348	0.015681	24.47987	1.268	0; 13.Oxidation(M);23.PFIND_DELTA_252.12;	3	spiP46013|KI67_HUN 2441 K.S/	1	0|0|PFIND_I	target	1	1.80229	96	252.124106
14	QE_Plus_Yangling_TCI	28336	2980.322	3		0 EGICALGGTSELSSEGTQHSYSEEEK		2980.319	0.003476	21.01752	1.305	0; 4.PFIND_DELTA_252.13;	3	spiP13797|PLST_HUI 100 K.V/	1	PFIND_D	target	0	-0.18048	36	252.121951
15	QE_Plus_Yangling_TCI	28826	2650.209	3		0 VVMALGDYMGASCHACIGGTNVR		2650.197	0.011875	27.98035	1.528	0; 9.Oxidation(M);13.Carbamidomethy(EC);1	3	spiP60842|IF4A1_HU 118 K.A/	1	0|0|0|PFRI	target	0	2.47523	96	252.121628
16	QE_Plus_Yangling_TCI	26229	3434.505	3		0 SSSQPSSCCSDPSKPGGNVEGATQSLAEQ		3434.507	-0.00218	26.26885	1.855	0; 9.PFIND_DELTA_309.15;	3	spiQ13501|SQSTM_I 281 K.K/	1	PFIND_D	target	1	0.18418	36	309.134533
17	QE_Plus_Yangling_TCI	23967	2451.174	2		0 VGVGTCGIADKPMTQYQDTSK		2451.172	0.002028	25.43208	2.145	0; 6.PFIND_DELTA_252.13;	3	spiO75940|SPF30_HI 208 K.V/	1	PFIND_D	target	1	-1.80833	32	252.122549
18	QE_Plus_Yl_FL_50per_	35837	4250.127	4		0 IIPTLEEGLQLPSPTATSQLPLESDAVECLN		4250.13	-0.00272	25.80814	2.425	0; 29.PFIND_DELTA_252.12;	3	spiP61978|HNRPK_H 103 K.G/	1	PFIND_D	target	0	-1.56726	36	252.100837
19	QE_Plus_Yangling_TCI	15312	2680.217	3		0 NGLGNQLSSGSHTSAASQCDSASSR		2680.205	0.012494	21.11898	3.606	0; 19.PFIND_DELTA_258.14;	3	spiQ5VV42|CDKAL_H 519 R.M/	1	PFIND_D	target	0	-1.08752	96	258.142127
20	QE_Plus_Yangling_TCI	36611	2789.418	3		0 LEVDANVAANSDLLGGGGVDGGHR		2789.398	0.02055	22.67217	4.388	0; 23.PFIND_DELTA_252.12;	3	spiQ9BQ69|MACD1_I 163 K.A/	1	PFIND_D	target	0	0.97621	100	252.129757
21	QE_Plus_Yangling_TCI	29513	2650.211	3		0 VVMALGDYMGASCHACIGGTNVR		2650.197	0.013714	25.2873	4.808	0; 3.Oxidation(M);13.Carbamidomethy(EC);1	3	spiP60842|IF4A1_HU 118 K.A/	1	0|0|0|PFRI	target	0	2.5457	96	252.123467
22	QE_Plus_Yangling_TCI	29511	2656.23	3		0 VVMALGDYMGASCHACIGGTNVR		2656.217	0.012958	25.26673	6.056	0; 3.Oxidation(M);13.Carbamidomethy(EC);1	3	spiP60842|IF4A1_HU 118 K.A/	1	0|0|PFRI	target	0	2.56769	96	258.142688
23	QE_Plus_Yangling_TCI	35823	3190.508	3		0 SGQAAKVDMVPGKPNCVESFSDYPPLGR		3190.508	-0.0001	24.22125	6.285	0; 16.PFIND_DELTA_252.13;	3	spiP68104|EF1A1_HI 395 K.F/	1	PFIND_D	target	1	-0.03766	36	252.11756
24	QE_Plus_Yangling_TCI	26068	2242.098	2		0 IHQVFHGSCITEGNELTK		2242.09	0.008404	27.60092	7.488	0; 9.PFIND_DELTA_252.12;	3	spiP78527|PRKDC_H 1895 K.T/	1	PFIND_D	target	0	1.49560	100	252.119734
25	QE_Plus_Yl_FL_50per_	32471	2859.302	3		0 PSWADQVEEEGEDDKCVTSELLK		2859.297	0.004956	25.85762	7.645	0; 16.PFIND_DELTA_252.12;	3	spiO75821|EIF3G_HI 9 K.G/	1	PFIND_D	target	1	-1.76907	32	252.1139
26	QE_Plus_Yl_FL_50per_	22104	3108.404	3		0 XEGICALGGTSELSSEGTQHSYSEEEK		3108.404	0.00005	34.2139	8.208	0; 5.PFIND_DELTA_252.12;	3	spiP13797|PLST_HU 99 R.Y/	1	PFIND_D	target	1	-2.73906	36	252.108028
27	QE_Plus_Yangling_TCI	33938	2088.077	2		0 KPANDITSQLENRGDLGR		2088.072	0.004788	24.23487	9.155	0;	3	spiQ8NC51|PAIRB_H 345 R.P/	1		target	1	0.4989	0	
28	QE_Plus_Yl_FL_50per_	25688	2345.159	2		0 KPTDGASSSNCVTDISHLVR		2345.158	0.000861	30.14683	1.055	0; 11.PFIND_DELTA_258.14;	3	spiP49321|NASP_HU 697 R.K/	1	PFIND_D	target	1	-1.46101	32	258.13179
29	QE_Plus_Yangling_TCI	26392	2996.312	3		0 EGICALGGTSELSSEGTQHSYSEEEK		2996.309	0.003507	19.14234	1.166	0; 4.PFIND_DELTA_268.12;	3	spiP13797|PLST_HU 100 K.V/	1	PFIND_D	target	0	1.73435	36	268.111921
30	QE_Plus_Yangling_TCI	23718	3209.427	3		0 AGEPNGPQAEEAHSPDVTAGCDPAGVHP		3209.416	0.010716	22.30986	1.205	0; 21.PFIND_DELTA_252.12;	3	spiQ08123|SUN12_H 737 R.-/	1	PFIND_D	target	0	3.35816	100	252.116304
31	QE_Plus_Yl_FL_50per_	4465	1783.745	2		0 GGGGGGGGGGGGCSGSGSSASR		1783.744	0.000579	27.09277	1.288	0; 12.PFIND_DELTA_258.14;	3	spiP42356|PAK4_HI 7 R.G/	1	PFIND_D	target	0	-1.47209	32	258.133684
32	QE_Plus_Yangling_TCI	25499	1954.071	2		0 VAPEEHPVLLTEAPLNPK		1954.064	0.007161	24.01486	1.325	0;	3	spiP60709|ACTB_HU 95 R.A/	1		target	0	1.43206	64	
33	QE_Plus_Yangling_TCI	35345	2919.447	3		0 LTTPTYGDLNHLVSATMSGVTTCLR		2919.432	0.015846	26.16512	1.326	0; 17.Oxidation(M);23.PFIND_DELTA_252.12;	3	spiP07437|TBB5_HU 216 K.F/	1	0|PFIND_I	target	0	1.28033	100	252.124552
34	QE_Plus_Yangling_TCI	22257	2330.104	3		0 ISCGGGWVDVVKQHPDGGPDGK		2330.102	0.002161	18.24762	1.328	0; 3.PFIND_DELTA_252.13;	3	spiQ86P45|DS03_H 283 R.A/	1	PFIND_D	target	0	-1.86942	32	252.12315
35	QE_Plus_Yangling_TCI	31201	2493.195	2		0 QGEYGLASICNGGGGASAMLIQK		2493.194	0.001113	25.01269	1.358	0; 10.PFIND_DELTA_252.13;19.Oxidation(M);	3	spiP24752|THIL_HU 403 K.L/	1	PFIND_D	target	0	0.68453	36	252.12147
36	QE_Plus_Yangling_TCI	29491	2702.278	2		0 VMTPYQPMPASSPVICAGGQDR		2702.271	0.006783	24.2682	1.385	0; 2.Oxidation(M);9.Oxidation(M);17.PFIND_	3	spiQ15365|PCBP1_H 177 R.C/	1	0|0|PFIND_I	target	0	1.92746	36	252.116332
37	QE_Plus_Yangling_TCI	24400	3107.48	3		0 GTEAGQVGEPGIPTGEAGPSCSSASDKLPI		3107.467	0.012457	21.53915	1.388	0; 21.PFIND_DELTA_252.12;	3	spiO15355|PPM1G_H 220 R.V/	1	PFIND_D	target	1	1.33006	100	252.120436
38	QE_Plus_Yangling_TCI	28553	2789.301	3		0 SPDEALPGGLSGCSSGSGHSPSYALER		2789.297	0.004013	19.655	1.388	0; 13.PFIND_DELTA_258.15;	3	spiQ5JZ5|PRC2B_H 1469 R.-/	1	PFIND_D	target	0	0.94009	36	258.143227
39	QE_Plus_Yangling_TCI	33407	2594.17	2		0 VCEDLDTSVNLAWTSGTNCTR		2594.169	0.001555	23.26515	1.415	0; 2.PFIND_DELTA_252.13;19.Carbamidome	3	spiP45880|VDAC2_H 208 K.F/	1	PFIND_D	target	0	-2.80033	32	252.121526
40	QE_Plus_Yangling_TCI	4684	1777.727	2		0 GGGGGGGGGGGGCSGSGSSASR		1777.724	0.002579	22.88986	1.566	0; 12.PFIND_DELTA_252.12;	3	spiP42356|PAK4_HI 7 R.G/	1	PFIND_D	target	0	1.18012	36	252.115707
41	QE_Plus_Yl_FL_50per_	35971	3190.51	3		0 SGQAAKVDMVPGKPNCVESFSDYPPLGR		3190.508	0.00133	23.51976	1.566	0; 16.PFIND_DELTA_252.13;	3	spiP68104|EF1A1_HI 395 K.F/	1	PFIND_D	target	1	-1.18682	36	252.11899
42	QE_Plus_Yangling_TCI	29398	2995.364	3		0 SYDPPCPGHWTPEAPGSTTCPGLPR		2995.353	0.010671	24.18949	1.655	0; 6.Carbamidomethy(EC);21.PFIND_DELTA_	3	spiO60936|NOL3_HI 103 R.A/	1	0|PFIND_I	target	0	0.75267	100	258.139094
43	QE_Plus_Yl_FL_50per_	25613	2339.139	2		0 KPTDGASSSNCVTDISHLVR		2339.138	0.00085	28.67104	1.705	0; 11.PFIND_DELTA_252.12;	3	spiP49321|NASP_HU 697 R.K/	1	PFIND_D	target	1	-1.33616	32	252.111802
44	QE_Plus_Yl_FL_50per_	25926	2980.31	3		0 EGICALGGTSELSSEGTQHSYSEEEK		2980.309	0.001464	20.55071	1.795	0; 4.PFIND_DELTA_252.12;	3	spiP13797|PLST_HU 100 K.V/	1	PFIND_D	target	0	-2.64583	36	252.109939

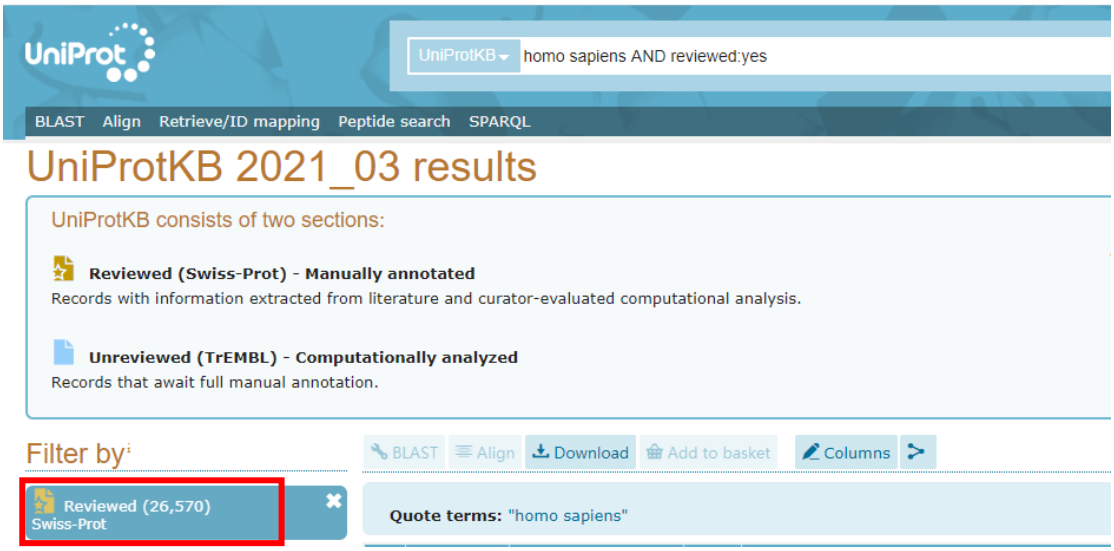
6. Supporting protocol 1: Protein sequence database

This protocol is used to download protein *.fasta files for database search.

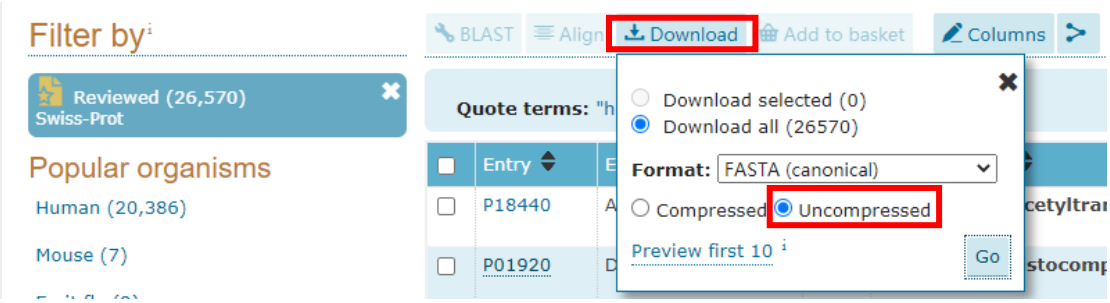
1) Open https://www.uniprot.org/, enter the Latin name of the species (e.g., *home sapiens*), then click search.



2) Click "*Reviewed*" (Swiss-Prot).



3) Select "*Uncompressed*", then Click "*Download*" and "*Go*".

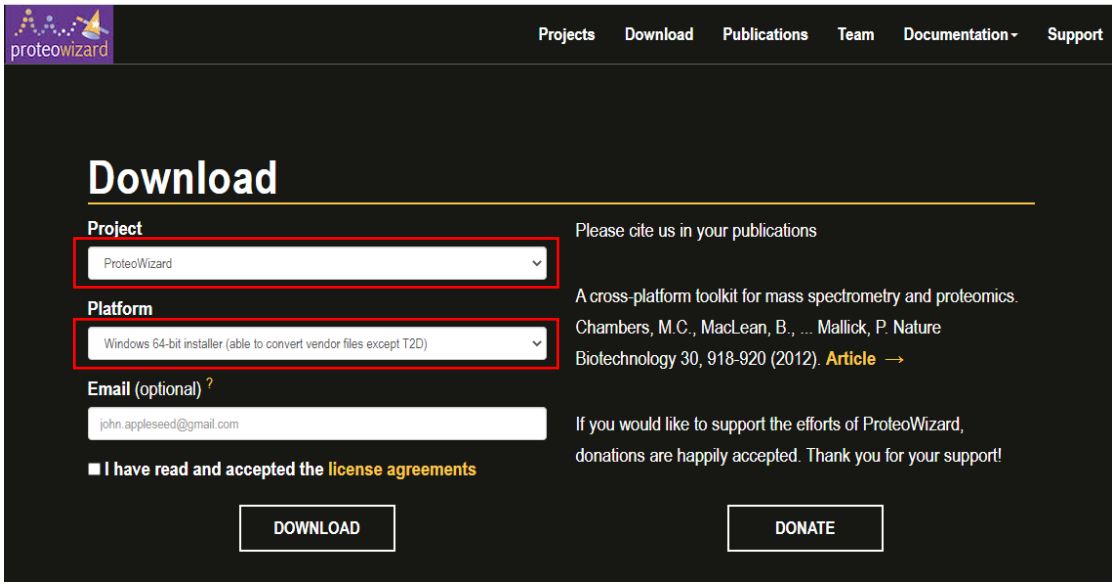


4) Get the *.fasta file.

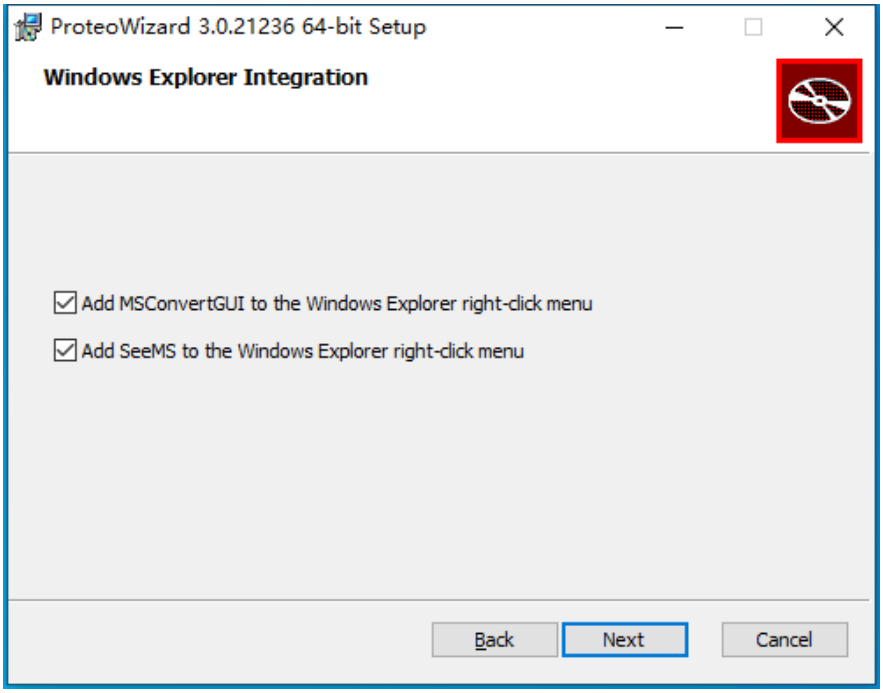
7. Supporting protocol 2: MSconvert

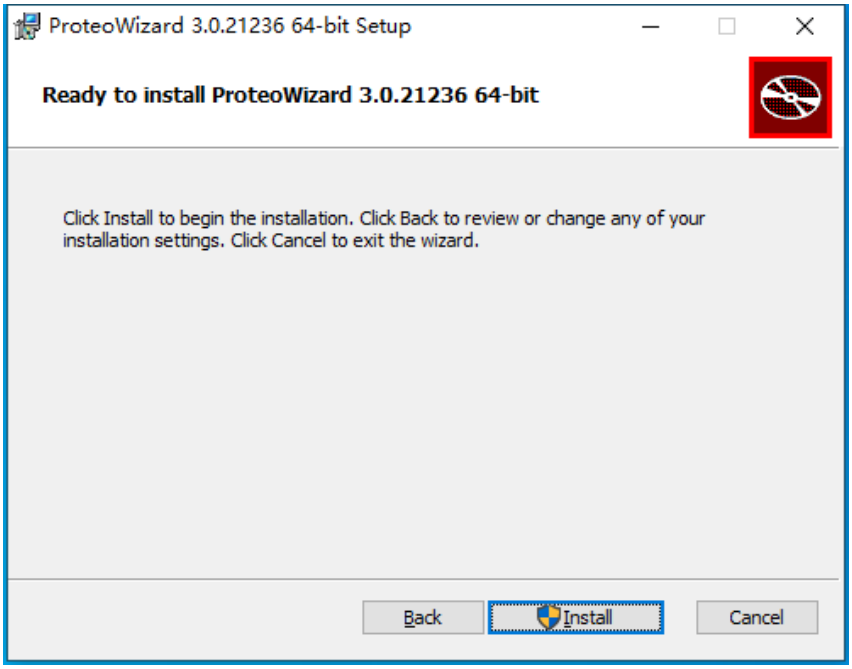
This protocol is used to convert non-Thermo MS data into mzML format files for pChem search.

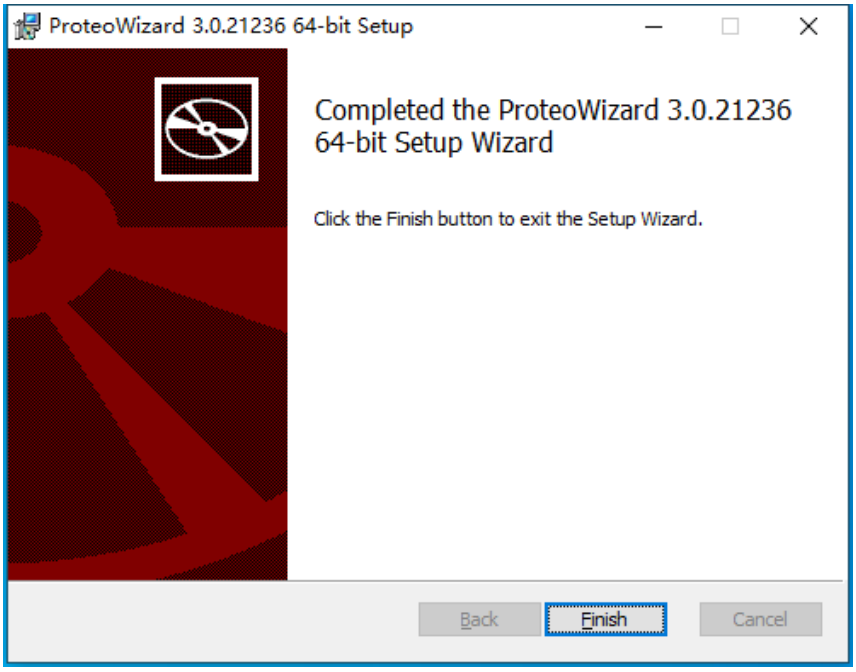
1) Download MSconvertGUI that is embedded in the ProteoWizard platform from: **https://proteowizard.sourceforge.io/download.html** .



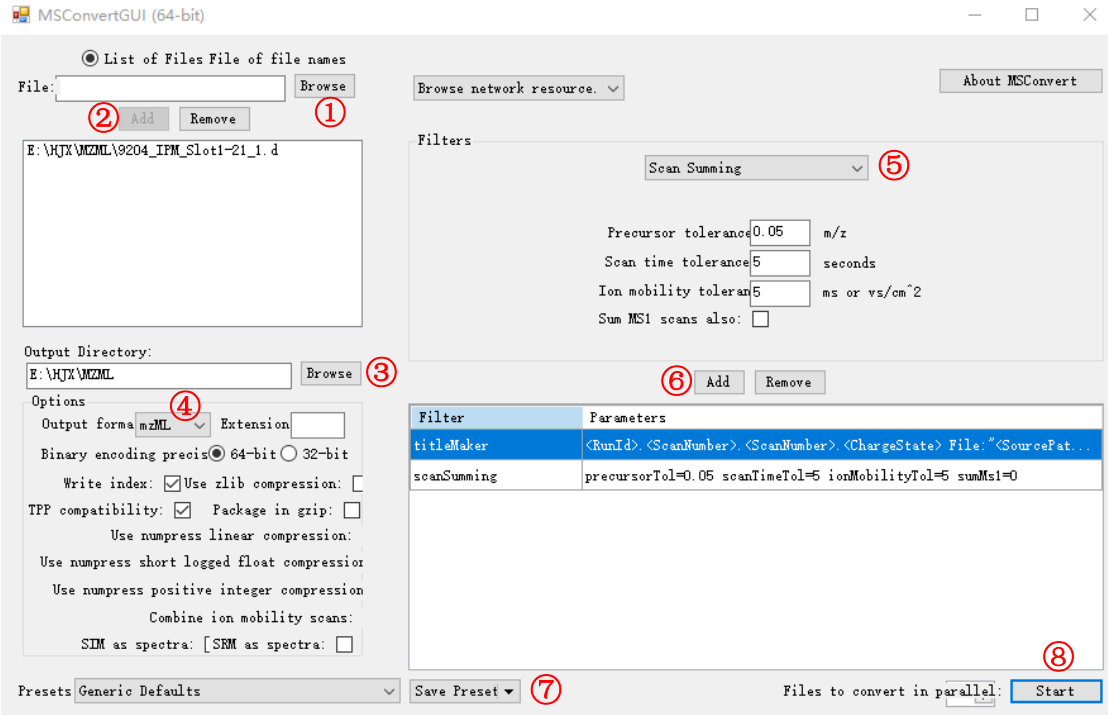
2) Install ProteoWizard according to the following instruction.




ProteoWizard 3.0.21236 64-bit Setup
Ready to install ProteoWizard 3.0.21236 64-bit
Click Install to begin the installation. Click Back to review or change any of your installation settings. Click Cancel to exit the wizard.
Back
Install
Cancel


ProteoWizard 3.0.21236 64-bit Setup
Completed the ProteoWizard 3.0.21236 64-bit Setup Wizard
Click the Finish button to exit the Setup Wizard.
Back
Finish
Cancel

3) Open MSconvertGUI



①-②Browse and add MS data (e.g., *.d, *.WIFF files)

③ Define output route

④ Choose *.mzML as the output data format

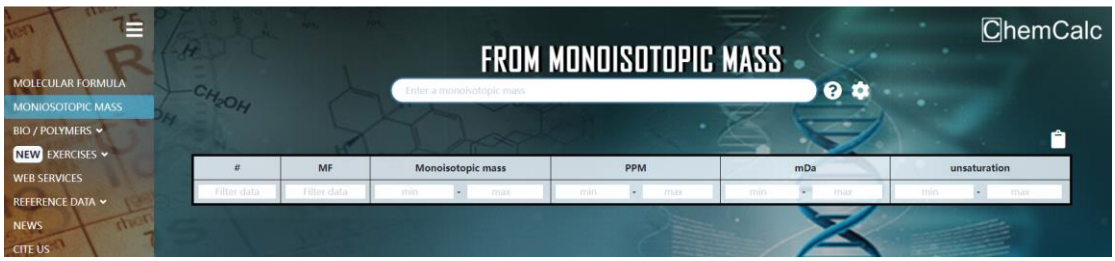
⑤ -⑥ Define parameters for Scan Summing

⑥ -⑧ Save and run

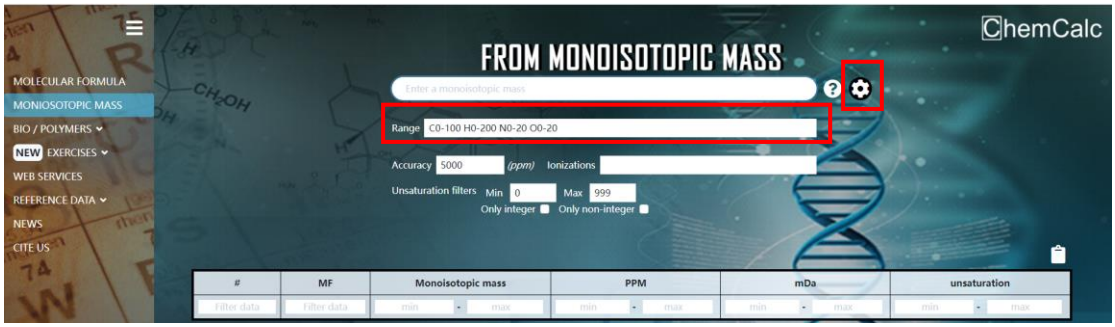
8. Supporting protocol 3: ChemCalc

This protocol is used to estimate candidate molecular formulas from the pChem-determined accurate masses.

1) Open **https://www.chemcalc.org/mf-finder** .



2) Click  , check the element composition.



3) Input the monoisotopic mass of each PDM shown in *pChem.summary* file. The candidate molecular formulas will immediately appear below.

Rank	PDM	Accurate Mass	Top1 Site|Probability	Others	#PSM	#PSM L|H	DFLs
1	PFIND_DELTA_252	252.122339	C|0.988		13876	7368|6508